Three Essays in Behavioral & Experimental Economics

Ben Leo Grodeck
Bachelor of Arts (Honours) & Commerce (Honours)

A thesis submitted for the degree of *Doctor of Philosophy* at
Monash University in *2023*
Department of Economics

Copyright notice

Abstract

This dissertation explores how psychological factors impact decision making and market efficiency within different domains. It consists of three self-contained chapters that use experimental economics methodology to investigate the role that these psychological factors play. The present thesis advances the current body of knowledge by providing novel insights into how cooperation and pro-sociality can be increased or decreased by psychological mechanisms and how it can either increase or constrain market efficiency. Specifically, it provides new insights into designing psychological mechanisms in buyer-seller markets, understanding of how compassion fade operates, and finally why people find certain harmless transactions repugnant.

In Chapter 1 we design and test a novel insurance advice mechanism aimed at promoting trust and cooperation in markets with asymmetric information. The simple mechanism we introduce is to have sellers advise buyers whether they should purchase third-party insurance. The theoretical model suggests that both cooperative and strategic sellers advise buyers not to purchase insurance. Once this advice has been given, strategic sellers are less likely to pursue self-interest due to associated psychological costs. We conduct a controlled laboratory experiment and show that the insurance advice mechanism significantly increases market efficiency, with sellers being more likely to cooperate with buyers and buyers being more likely to purchase from sellers.

Chapter 2 shifts the focus from market efficiency to the domain of pro-social behavior. In this chapter we investigate the phenomenon of compassion fade: a tendency to act less altruistically when faced with more, rather than fewer individuals in need. Using variations of the dictator game in both lab and online experiments, our design allows us to explore both the determinants of compassion fade, and the mechanism by which it operates. We find that adding a second unhelpable individual who is in a needy state significantly increases the rate of selfish behavior. However, when the unhelpable recipient is not in a needy state, decision makers act significantly less selfishly, compared to when they are in a needy state. Finally, we present evidence that the mechanism of compassion fade is diminishing negative affect (e.g., guilt) that arises from selfish decisions, rather than through the diminishment of warm glow.

Chapter 3 studies whether and why people feel repugnance towards *harmless* transactions that profit off others' misfortune, without causing the misfortune. In a series of online experiments that vary in the moral intensity of misfortune—from monetary losses in a game to deaths from road accidents—we find robust evidence of repugnance, measured using costly second- and third-party punishment, towards the party who profits from others' negative outcomes (that are merely determined by luck). Overall, we find that repugnance is mainly outcome-based: people dislike profit-making that occurs as a result of others' (mis)fortune.

Finally, this thesis concludes by summarizing the main findings from the three chapters, the policy implications, and what future research questions arise as a result of this dissertation.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes zero original papers published in peer reviewed journals and two submitted publications. The core theme of the thesis is how psychological factors impact decision making and market efficiency. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within The Department of Economics under the supervision of Professor Erte Xiao, Professor Lata Gangadharan, and Professor Philip J. Grossman.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

In the case of *chapters 1, 2 and 3* my contribution to the work involved the following:

Thesis Chapter	**Publication Title**	**Status** *(published, in press, accepted or returned for revision, submitted)*	**Nature and % of student contribution**	**Co-author name(s) Nature and % of Co-author's contribution***	**Co-author(s), Monash student Y/N***
1	*To Insure or Not to Insure? Promoting Trust and Cooperation with Insurance Advice in Markets*	*Returned for Revision*	*50%. Concept, experimental design, running the experiment, data analysis, writing and editing the paper.*	*1) Erte Xiao, input into manuscript 30%* *2) Chensi Wang, input into manuscript 15%* *3) Franziska input into manuscript 5%*	*No* *No* *No*
2	The Effect of Compassion Fade on Altruistic Behavior: Experimental Evidence for a Guilt Mitigation Account	Under Review	*60%. Concept, experimental design, running the experiment, data analysis, writing and editing the paper.*	*1) Toby Handfield, input into manuscript 30%* *2) Matthew Kopec, input into manuscript 10%*	*No* *No*
3	No (Intention to) Profit; No Repugnance? Evidence From Online Experiments.	Submitted	*50%. Concept, experimental design, running the experiment, data analysis, writing and editing the paper.*	*1) Erte Xiao, input into manuscript 20%* *2) Nina Xue, input into manuscript 30%*	*No* *Yes*

I ~~have~~ / have not renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Acknowledgements

I would not be sitting here today writing this Acknowledgements section without the help, support, and mentorship from many people. Back in 2017, I was trying to decide whether to go overseas or stay here at Monash for my PhD. While sometimes I do think about the counterfactual, I have never regretted my decision to stay here. In fact, it's probably the best decision I have ever made. I am truly thankful for my friends, family, and mentors for making that a great decision.

First, I would like to thank my PhD supervisors, Erte Xiao, Lata Gangadharan, and Philip J. Grossman. To Lata and Phil, I consider myself extremely lucky to have both of you as mentors. Your body of research has always inspired me, and I have learnt so much from both of you. Your support has been invaluable, always taking the time to help me when I needed it despite your busy schedules. Thank you for everything.

To Erte, it was almost six years ago when we met for the first time. While I was looking for an honours' advisor, we had a meeting where I talked about my weird, wacky, and outlandish research ideas. Not only were you willing to listen and take me seriously, but you also encouraged me to pursue these interests and investigate the questions I was interested in. Since that moment, you have been the greatest mentor one could ask for. You have supported me through thick and thin, always willing to go into battle for me, and I appreciate you pushing me to be the best scholar and person I can be. I always enjoy our conversations about research projects and the fascinating ideas we come up with together. I hope our working relationship can continue as colleagues in the future. Erte, I would not be here today without your mentorship. Thank you for everything.

To Toby Handfield, my unofficial fourth advisor. You have been so generous with your help throughout the years and thanks for always believing in me. I'm so lucky to have you as a mentor, a colleague, and a friend.

Second, I want to thank my colleagues and other academics for the incredible support I have received. To the Monash BET community and faculty in general—Birendra, Choon, Klaus, Vai-Lam, Andreas, Xiaojian, Chengsi. Thank you for the inclusive environment you created. I never felt like just a PhD student around you, but rather a colleague.

I've also been very fortunate to meet incredible scholars who have helped and supported me, even though they had no obligation to do so. I want to thank Kirby Nielsen, Marta Serra-Garcia, Oliver Houser, and David Reinstein for their mentorship and kindness.

I would also like to take a moment to thank the amazing friends I have made throughout my PhD process. To David and Nina, I was very fortunate to have my PhD overlap with both of you. You're not just my colleagues, but also my friends. To Zach, and Philipp. I am so happy I got the chance to work with you, learn from you, and become friends with you. I hope we can continue to build these friendships in the future. To Dave, thanks for putting up with me as a housemate for four months in Oxford.

To The Global Priorities Institute and The Forethought Foundation. Thanks for believing in me—a PhD student from a lesser-known university—even when I didn't believe in myself. Your help and support have been invaluable. I am forever grateful for the opportunities you have given me.

Third, I want to thank my friends here in Melbourne. To 'The Peanuts', 'Friends who Activitise', Shorty, and Elliot. Thanks for always being there for me. I am seriously lucky to have so many amazing friends.

I wanted to specifically thank my best friend Gabriela D'Souza, who has been my rock and confidant these last 5 years. I am so lucky to have a kind-hearted person, such as yourself in my life. You're not just a friend, you're family. Thanks for helping me through all the tough times.

Fourth, I wanted to thank my incredible family. To my parents, Lisa and Anton. You've always been my greatest supporters. I truly appreciate everything you have done for me these past 32 years. I love you both so much and I hope you get some nachas from me completing the dissertation! Thank you for everything. To Adam, we have such a strong brotherly bond. Thanks for always having my back. To Aunty Fi. Thanks for always opening your home to me in Sydney and always being there for me to talk to.

Finally, I want to dedicate this thesis to my Uncle David Philips who passed away in 2020. David was my number one fan. We shared an incredibly close bond that only got stronger with time. Without fail, he would always be there for me no matter what. Words cannot describe how much I miss you. I know how proud you would be right now.

Table of Contents

Introduction

One of the key contributions of the behavioral economics literature is enriching traditional economic models that explain behavior, without implying inconsistent preferences. This is achieved by incorporating different psychological factors into these utility functions, such as inequality aversion (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), beliefs (Bénabou and Tirole, 2016), guilt (Charness and Dufwenberg, 2006), and warm glow (Andreoni, 1989, 1990). This has allowed economists to gain a deeper understanding of people's preferences and subsequent behavior, resulting in greater explanatory and predictive power of economic theory. It has also added to the toolbox of economists, allowing them to identify market inefficiencies and design new mechanisms and interventions to solve it.

The present thesis contains three self-contained chapters. While the domain of interest does differ across each chapter, there are common themes present. First, in each chapter psychological factors are introduced into a theoretical model, generating different behavioral predictions. Second, all chapters use the methodology of experimental economics to investigate the research question of interest.

This thesis contributes to the literature by designing novel experiments and providing new evidence of how psychological factors can affect both individual behavior and market efficiency across three different domains. These domains are markets with asymmetric information, pro-social behavior, and repugnance. This introduction provides an overview of each chapter and how it relates to the common theme of the thesis. There is also a brief discussion about the methodology, and how it differs in each chapter.

Chapter 1 designs and tests a novel insurance advice mechanism aimed at promoting trust and cooperation in markets with asymmetric information. Previous research has proposed and tested innovative solutions to this problem, including the widely studied reputation mechanism (for a review, see Chen et al., 2021). However, reputation mechanisms are subject to o problems such as missing information (Resnick and Zeckhauser, 2002; Bolton et al., 2004; Dellarocas and Wood) and manipulating reviews (Mayzlin et al., 2014). Another well-known solution, third-party insurance, is often costly and comes with exclusions and limitations. Moreover, insurance provided by a third party may not change the incentives for sellers or manufacturers to cooperate and thus have a limited impact on improving consumers' willingness to trade. The proposed mechanism builds upon third-party insurance, by having sellers advise buyers whether to purchase third-party insurance against the potential losses from the opportunistic behavior of strategic sellers. The theoretical model suggests that both cooperative and strategic sellers advise buyers not to purchase insurance. Once this advice has been given, strategic sellers are less likely to pursue self-interest due to associated psychological costs. A controlled laboratory experiment is conduced and shows that the insurance advice mechanism significantly increases market efficiency, with sellers being more likely to cooperate with buyers and buyers being more likely to purchase from sellers. This chapter highlights how understanding psychological factors can be used to design mechanisms that increase market efficiency.

Chapter 2 investigates the phenomenon of compassion fade: a tendency to act less altruistically when faced with more, rather than fewer individuals in need, even if the options available to the decision maker are the same (Butts et al., 2019; Erlandsson et al., 2014; Markowitz et al., 2013; Small et al., 2007; Västfjäll et al., 2014). Although the phenomenon of compassion fade has been well documented in the psychology literature, there are several limitations of this existing research program. This chapter contributes to this literature in several ways. First, it distinguishes two components of compassion fade, the number of unhelpable individuals in the decision frame and the degree of need of those unhelpable individuals. Second, it attempts to quantify compassion fade. Third, it investigates whether the compassion fade phenomenon exists in a pro-social setting, as well as an altruistic setting. Finally, it provides evidence for how the mechanism by which compassion fade operates. Whether it is due to shifting normative standards, diminished disutility ("guilt") for selfish decisions or diminished warm glow for generous decisions.

Using variations of the dictator game, the experiment is designed specifically to explore both the determinants of compassion fade, and the mechanism by which it operates. Two experiments are conducted, one in the laboratory and a follow up experiment on Amazon mTurk. The main finding is that compassion fade does exist in a pro-social context. Adding a second unhelpable individual who is in a needy state significantly increases the rate of selfish behavior. However, compassion fade is sensitive to the level of "need" – of the additional individual. When the unhelpable recipient is not in a needy state, decision makers act significantly less selfishly, compared to when they are in a needy state. Finally, there is evidence that the mechanism of compassion fade is diminishing negative affect (e.g., guilt) that arises from selfish decisions, rather than through the diminishment of warm glow. Overall, this chapter takes insights from psychology and uses the tools of behavioral and experimental economics to generate a motivating framework for how compassion fade operates, which generates hypotheses that are tested using rigorous experiments.

Finally, **Chapter 3** explores how and why people feel repugnance towards harmless transactions that profit off others' misfortune, without causing the misfortune. Transactions perceived as repugnant often face significant constraints and can even be legally banned (Roth, 2007; Elias et al., 2017). Our understanding of the underlying drivers of the repugnant feelings, however, remains limited. One frequently cited explanation is the potential harm imposed by the profiting activities on others, such as exploitation and risk to the buyers or sellers (Satz, 2010; Sandel, 2012; Leuker et al., 2021). Yet, it is unclear whether harm is a necessary condition for repugnance. While these transactions do not cause the negative events, a common feature is that profits are attached to the occurrence of others' misfortunes. The chapter defines this phenomenon "*piggyback profiting*", whereby one party simply profits from the outcomes of another, without having any role in the outcome itself. This paper takes a first step to investigate whether people feel repugnance towards harmless piggyback profiting transactions and potential factors that could explain why people might find such transactions repugnant.

Through a series of online experiments that vary in the moral intensity of misfortune—from monetary losses in a game to deaths from road accidents—there is robust evidence of repugnance, but intention to profit off others' misfortune only plays a limited role in punishment decisions. robust evidence of repugnance, measured using costly second- and third-party punishment, towards the party who profits from others' negative outcomes (that are merely determined by luck). Intentions to profit from others' misfortune affect the punishment decisions of second parties but not third parties. Repugnance is observed even when the profits are associated with good outcomes. Overall, repugnance is mainly outcome-based: people dislike profit-making that occurs as a result of others' (mis)fortune.

Experimental Methods and Subject Pools

This thesis highlights how different subject pools can be used to conduct economic experiments. The taxonomy of Harrison and List (2004) divides experiments into four different categories. Two of these categories, conventional lab experiments and artefactual field experiments are used across the three chapters. Harrison and List (2004) define Artefactual field experiments as lab experiments run with a non-standard subject pool (non-students). The opportunities of experiments of this kind have increased over recent years with the formation of online experimental platforms, such as Amazon mTurk and Prolific. These platforms have already received significant uptake in experimental economics (Exley & Kessler, 2019; Exley, 2019; Hauser & Schwarz, 2016; Serra-Garcia & Szech, 2019).

There are tradeoffs between moving from the lab to an online platform. The benefit of online platforms is that it is cheaper to recruit subjects, allowing larger sample sizes and consequently higher-powered studies. Furthermore, as Gandullia et al., (2020, p. 2) have argued, moving from a university student sample to an online sample may also reduce experimenter demand effects as the experimenters are not physically present at the time of data collection thus further making plausible this choice of participant recruitment.

However, there are still concerns about the control experimenters have on these platforms. First, can the experimenter ensure that they have the participant's full attention throughout the experiment and properly understand the tasks? Second, since researchers from other disciplines (such as psychology) use these platforms, participants in the subject pool may have experienced or expect deception. Regarding noise, Gupta et al., (2021) find that "noisy behavior accounts for 60% of the observations on mTurk, 19% on Prolific, and 14% for the lab". Despite the noise on mTurk, due to how cheap each observation is, it still can have "greater inferential power than the laboratory." While both online platforms do well compared to the lab in terms of accounting for noise, they are less elastic in terms of responding to interventions (smaller effect sizes). Ultimately, this means that they can have less power compared to laboratory studies.

In Chapter 1 a traditional laboratory experiment is used. The experiment is dynamic in the sense that buyers and sellers interact in real time, and multiple rounds of the game are used. The experiment uses stranger matching between buyers and sellers. Logistically, this experimental design would be very difficult to run on an online platform, such as Prolific. If any participants drop out during a session, this would be problematic. Hence, a lab experiment was appropriate for this study. In Chapter 2 both a lab experiment and an online experiment are used. The key reason for using the online experiment was to increase the sample size of observations. Having a within subject design (multiple tasks) meant that order effects may be a problem. By having approximately n=720 observations, this allowed almost all 720 orders of the tasks to help control for this. The experiment was conducted on mTurk using Cloud Research (Litman et al., 2007), which was used to help control for data quality. Finally, Chapter 3 also used an online experiment, this time using Prolific. There were two key reasons for using Prolific. First, once again to increase the sample size of the experiment, given one of the interests was punishment on the intensive margin, and observations would be smaller given not every participant chooses to punish. Second, studying repugnance using only a student sample may be limited due to external validity. By studying the behavior of a broader population (US citizens) it may provide a more generalizable understanding of repugnance.

This thesis illustrates the flexibility that experimental economists now have with the emergence of online platforms, when it comes to designing and running experiments.

References

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and Ricardian equivalence. *Journal of political Economy*, *97*(6), 1447-1458

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The economic journal*, *100*(401), 464-477

Bénabou, R., & Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives, 30*(3), 141-64

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American economic review*, *90*(1), 166-193

Butts, M. M., Lunt, D. C., Freling, T. L., & Gabriel, A. S. (2019). Helping one or helping many? A theoretical integration and meta-analytic review of the compassion fade literature. *Organizational Behavior and Human Decision Processes*, *151*, 16–33. https://doi.org/10.1016/j.obhdp.2018.12.006

Bolton, G. E., Katok, E., and Ockenfels, A. (2004). How effective are electronic reputation mechanisms? An experimental investigation. *Management Science*, *50*(11), 1587-1602

Charness, G., & Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, *74*(6), 1579-1601

Chen, Y., Cramton, P., List, J. A., and Ockenfels, A. (2021). *Market Design, Human Behavior, and Management*. Management Science. 67(9). 5317-5348

Dellarocas, C., and Wood, C. A. (2008). The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Science*, *54*(3), 460-476
Elias, J., Lacetera, N., Macis, M., & Salardi, P. (2017). Economic Development and the Regulation of Morally Contentious Activities. *American Economic Review*, *107*(5), 76–80

Erlandsson, A., Björklund, F., & Bäckström, M. (2014). Perceived Utility (not Sympathy) Mediates the Proportion Dominance Effect in Helping Decisions: Perceived Utility Mediates the PDE. *Journal of Behavioral Decision Making*, *27*(1), 37–47

Exley, C., & Kessler, J. (2019). *Motivated Errors* (No. w26595; p. w26595). National Bureau of Economic Research

Exley, C. L. (2019). Using charity performance metrics as an excuse not to give. *Management Science*, mnsc.2018.3268. https://doi.org/10.1287/mnsc.2018.3268

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, *114*(3), 817-868

Gandullia, L., Lezzi, E., & Parciasepe, P. (2020). Replication with MTurk of the experimental design by Gangadharan, Grossman, Jones & Leister (2018): Charitable giving across donor types. *Journal of Economic Psychology, 78,* 102268

Gupta, N., Rigotti, L., & Wilson, A. (2021). The Experimenters' Dilemma: Inferential Preferences over Populations. *arXiv preprint arXiv:2107.05064*

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic literature*, 42(4), 1009-1055

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, 48(1), 400–407

Leuker, C., Samartzidis, L., & Hertwig, R. (2021). What makes a market transaction morally repugnant? *Cognition*, *212*, 104644

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Markowitz, E., Slovic, P., Vastfjall, D., & Hodges, S. (2013). *Compassion fade and the challenge of environmental conservation*. https://scholarsbank.uoregon.edu/xmlui/handle/1794/22102

Mayzlin, D., Dover, Y., and Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, *104*(8), 2421-55

Resnick, P., and Zeckhauser, R. (2002). Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system. *The Economics of the Internet and E-commerce*, 11(2), 23-25

Roth, A. E. (2007). Repugnance as a Constraint on Markets. *Journal of Economic Perspectives*, *21*(3), 37–58

Sandel, M. J. (2012). *What money can't buy: The moral limits of markets*. Macmillan

Satz, D. (2010). *Why some things should not be for sale: The moral limits of markets*. Oxford University Press

Serra-Garcia, M., & Szech, N. (2022). The (in) elasticity of moral ignorance. *Management Science*, *68*(7), 4815-4834

Small, D. A., Loewenstein, G., & Slovic, P. (2007). Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behavior and Human Decision Processes*, *102*(2), 143–153

Västfjäll, D., Slovic, P., Mayorga, M., & Peters, E. (2014). Compassion Fade: Affect and Charity Are Greatest for a Single Child in Need. *PLOS ONE*, *9*(6), e100115

Chapter 1: To Insure or Not to Insure? Promoting Trust and Cooperation with Insurance Advice in Markets

Ben Grodeck, Franziska Tausch, Chengsi Wang, and Erte Xiao*

Abstract: We design and test a novel insurance advice mechanism aimed at promoting trust and cooperation in markets with asymmetric information. In a buyer-seller game with third-party insurance, sellers have the option to advise buyers on whether to purchase insurance against the potential losses from the opportunistic behavior of strategic sellers. We hypothesize that advising not to purchase insurance introduces a psychological cost for defection. We develop a theoretical model that selects a pooling equilibrium where both cooperative and strategic sellers advise buyers not to purchase insurance. Once this advice has been given, strategic sellers choose not to defect if the associated psychological costs are sufficiently large. Data from a controlled laboratory experiment shows that the insurance advice mechanism significantly increases market efficiency, with buyers being more likely to purchase from sellers and sellers being more likely to cooperate. Furthermore, we find that the insurance advice mechanism is more effective when sellers can observe buyers' insurance purchase decisions.

JEL codes: C91, D9, D47, D82, L86

Keywords: asymmetric information, insurance, trust, cooperation, communication, experimental economics

Acknowledgments: The authors thank Arthur Campbell and Joshua Miller, as well as seminar and conference participants at the University of Queensland's BESC e-seminar, M-BEEs/M-BEPs 2019 conference, Virtual East Asia Experimental and Behavioral Economics Seminar, East China Normal University, and ANZWEE 2019 conference for valuable feedback and comments.

*Grodeck: Department of Economics, Monash University. Email: ben.grodeck1@monash.edu
Tausch: Stepstone. Email: FranziskaTausch@web.de
Wang: Department of Economics, Monash University. Email: chengsi.wang@monash.edu
Xiao: Department of Economics, Monash University. Email: erte.xiao@monash.edu

1. Introduction

Asymmetric information is ubiquitous in economic transactions. Consumers are often unable to verify sellers' credibility before purchasing a product. In this case, if the consumer does not trust the seller, they may refrain from the transaction entirely. Many mechanisms have been designed to solve this asymmetric information problem and facilitate efficient transactions. Among them, insurance (or warranties) is a common practice. In particular, the buyer can purchase insurance or warranties from a third-party provider to protect his purchase. For example, online markets such as eBay offer the option of purchasing warranties from the third-party provider Squaretrade. Buyers can purchase the warranty either at the time of buying the product on eBay or directly from Squaretrade's website after purchasing the product (Steiner, 2012). The insurance provided by a third party offers additional or extended coverage to the existing manufacturer's warranty or protection when consumers purchase products, especially second-hand ones, for which the manufacturer's warranty is not honored. However, such insurance is often costly and comes with exclusions and limitations. If buyers are unwilling to pay the cost of insurance or are discouraged by complicated exclusion clauses, the inclusion of insurance may not generate more transactions. Moreover, insurance provided by a third party may not change the incentives for sellers or manufacturers to cooperate and thus have a limited impact on improving consumers' willingness to trade. In this paper, we propose and test a novel insurance advice mechanism aimed at promoting trust and cooperation in markets with asymmetric information.

The key to our proposed mechanism is to allow the seller (the party who has more information) to advise the buyer whether he should purchase third-party insurance[1]. In general, insurance often addresses two different types of risks: 1) risks about the seller's cooperative type, such as her intention to deliver the product on time or her intention to produce a high-quality product as advertised; and 2) natural risks that are out of her control, such as bad weather that causes the delay of the shipment. As our focus is on the asymmetric information problem, the proposed advice mechanism is related to insurance against the first type of risk. Our mechanism builds on markets where an insurance option is already in place, such that the marginal cost of introducing the advice option is negligible. As a first step, we test the mechanism built on insurance provided by a third party instead of the seller. This

[1] For simplicity, we will use "she" to refer to the seller and "he" to refer to the buyer.

feature avoids potential confounds due to the additional profit incentives that sellers may have to sell insurance.

We hypothesize that advising not to purchase insurance introduces a psychological cost for defection. First, giving advice may lead the seller to feel more accountable for the buyer's payoffs, as she now plays a more active role in the buyer's decision (Tetlock, 1985; Lerner and Tetlock, 1994; 1999). If the seller is subject to omission bias (Ritov and Baron, 1992), she may judge defection—after advising the buyer not to protect himself from the risk—as morally worse than when she does not exert any influence on the buyer's decision. Second, if the advice of not purchasing insurance is taken as a statement that the seller will cooperate, subsequent defection may render the advice a lie and inflict psychological costs due to lying aversion (Cressey, 1986; Gneezy et al., 2013; Abeler et al., 2014; Abeler et al., 2019). Third, if advising not to purchase insurance increases the buyer's expectation that the seller will cooperate, the seller may be averse to disappointing the buyer (Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007; Balafoutas and Sutter, 2017; Cartwright, 2019). Finally, these above-mentioned psychological mechanisms may mitigate the moral wiggle room that sellers can exploit to pursue self-interest (Dana et al., 2007; Gino et al. 2016; Benabou and Tirole, 2016). For example, without insurance advice, the seller can plausibly justify her defection through reasoning, such as "It was the buyer's decision, I am not responsible for them choosing that." Such self-serving reasoning is no longer easy after the seller takes a more active role in the buyer's decision, or has to lie to defect, and thereby opportunistic behaviour becomes more psychologically costly.[2]

We consider a theoretical model that contains both cooperative sellers, who always ship the product, and strategic sellers, who ships the product only if doing so maximizes their utility. Our model predicts that a pooling equilibrium with both types of sellers advising not to purchase insurance is likely to emerge, provided psychological costs are sufficiently large. At equilibrium, the buyer follows the advice and purchases the product without insurance, and the seller subsequently ships the product. As a result, the insurance advice mechanism achieves more efficient trades.

We conduct a controlled laboratory experiment to examine the effectiveness of the mechanism empirically. In particular, we address two main research questions. Does the insurance advice mechanism increase the number of buyers who enter transactions with

[2] On the other hand, advising the buyer to purchase the insurance could reduce the seller's accountability for the buyer's losses and increase the moral wiggle room for excusing her opportunistic behaviour.

sellers? Are sellers more likely to cooperate with buyers under the insurance advice mechanism?

Although in online marketplaces such as Amazon and eBay, buyers' insurance purchase decisions can be easily made observable to sellers, we take into account the fact that sellers do not always observe the buyers' insurance purchase decisions when the insurance is provided by a third party. Theoretically, the insurance advice mechanism can help build trust and improve efficiency even if the seller does not observe the buyer's actual insurance purchase decision. This is the case because the seller expects the psychological cost to be associated with not shipping the product and becomes more likely to ship the product when she anticipates that the buyer may buy the product without insurance. In turn, knowing that the seller may ship the product, the buyer is also willing to follow the insurance advice with some non-zero probability. The improvement, however, is not as effective compared to when sellers perfectly observe the insurance purchase decisions. We empirically test whether the effectiveness of the mechanism varies based on the observability of the buyer's insurance purchase decision.

The experiment consisted of three treatments. The control treatment was a buyer-seller game with insurance. In the game, the buyer decided whether to purchase a product, and the seller decided whether to ship the product upon receiving the payment. If the buyer purchased the product, he could also purchase insurance against the risk that the seller might not ship the product after receiving the payment. To test the insurance advice mechanism, we designed two treatments: insurance advice (IA) treatment and insurance advice with hidden information (IA_HI) treatment. In both treatments, we introduced an insurance advice mechanism in which the seller had to advise the buyer whether to purchase the insurance before starting the buyer-seller game. Upon receiving the advice, the buyer decided whether to buy the product and, if so, whether to purchase the insurance. In the IA treatment, if the buyer purchased the product, the seller was informed of the buyer's insurance purchase decision before deciding whether to ship the product. In the IA_HI treatment, the seller never learned about the buyer's insurance purchase decision. This is the only difference between the two treatments. We used shipping as a simple way to introduce defections in the game. If the proposed mechanism works in this setting, it should also effectively reduce other types of defections, such as selling faulty products.

Our findings are consistent with our hypotheses. In the IA treatment, sellers advise not to purchase insurance 81% of the time. Compared with the control treatment, the rate of product purchases increases by approximately 31% in the IA treatment. Whereas buyers purchased the product 74% of the time when sellers advised not to purchase insurance, they purchased it only 35% of the time when the advice was to purchase insurance. The number of sellers who shipped the product also increases by almost 40%. These improvements increase the proportion of efficient trades and the average profit for both buyers and sellers. The mechanism remains effective in the IA_HI treatment. About 71% of sellers advised not to purchase insurance. Compared to the control, the product purchase rate increases by 33%, and the shipping rate increases by approximately 22% in the IA_HI treatment, increasing market efficiency. However, as our theoretical model predicts, compared to the IA treatment, the IA_HI treatment is less efficient because buyers were less likely to follow the advice of no insurance, and sellers were slightly less likely to ship the product.

This paper contributes to two strands of the literature. One is research on market design aimed at solving market failures due to information asymmetry. Several innovative solutions have been proposed and tested, including the widely studied reputation mechanism (for a review, see Chen et al., 2021). Previous studies have examined how to improve the reliability of reputation mechanisms that are subject to problems such as missing information (Resnick and Zeckhauser, 2002; Bolton et al., 2004; Dellarocas and Wood, 2008; Cabral and Hortacsu, 2010; Li and Xiao, 2014; Bolton et al., 2018; Bolton et al., 2019) and manipulating reviews (Mayzlin et al., 2014).

The simple mechanism we propose complements this literature by pointing out a new direction for solutions. For example, on eBay, buyers are offered extended warranties via Squaretrade or xcover.com. To implement the insurance advice mechanism, eBay could allow sellers to recommend to the buyer whether he should purchase the extended warranty. The advice mechanism can be especially beneficial for new sellers before they are able to establish a positive reputation via a feedback mechanism. The mechanism can also work in offline markets, such as the used car market. For instance, the original car owner may suggest whether the potential buyer should purchase an extended warranty from a third party.

Although our main interest is to propose and test the effectiveness of the insurance advice mechanism, it is interesting to consider how the advice mechanism relates to other types of communication (in particular promises) that have been shown to be effective in promoting cooperation (Ellingsen and Johannesson, 2004; Binmore, 2006; Charness and Dufwenberg, 2006; Bicchieri and Lev-On, 2007; Vanberg, 2008; Sanchez-Pages and Vorsatz, 2009; Erat and Gneezy, 2012; Battigalli et al., 2013; López-Pérez and Spiegelman, 2013). Sally (1995) conducted a meta-analysis and found that communication was the most effective factor in promoting cooperation in prisoner’s dilemma experiments. In particular, the analysis shows that elicited promises have a strong additional effect on cooperation. Recently, there has been emerging experimental research on the relationship between promises and cooperation. For example, researchers have shown that people follow their promise to either avoid lying costs (Ellingsen and Johannesson, 2004; Vanberg, 2008; 2013; Serra-Garcia et al., 2013) or as a result of guilt aversion (Charness and Dufwenberg, 2006; Battigalli et al., 2013). In our experiment, when the seller advises not to purchase insurance, the buyer may interpret this message as an implicit promise to ship the product.

It is worth noting that the literature on communication and promises highlights that compared to unconstructed free-form communication, restricted-form communication, such as binary messages, is often much less or not effective at all (Bracht and Feltovich, 2009). For example, a pre-formulated promise, such as "I promise to cooperate," may have a smaller effect than free-form communication or no effect at all on increasing cooperative behavior, specifically in strategic environments (Lundquist et al., 2009; Charness and Dufwenberg, 2010; Belot et al., 2010; Chen and Zhang, 2021; Brandts et al., 2019). The advice in our setup is given in a binary form. Sellers only choose between "Advise the buyer to purchase the insurance" or "Advise the buyer not to purchase the insurance." Thus, if the buyer perceives the advice as an implicit promise, it is, at best, a weak and indirect bare promise. The significant effect of a plain insurance advice message suggests that there may be some fundamental differences between advice and bare promise and that additional psychological channels may play a role.

According to the philosophy of language (Searle, 1975), the advice mechanism theoretically differs from bare promises. When a seller utters a bare promise to cooperate, she 'commits' to cooperate (without explicitly asking the buyer to take a certain action). This promise can be seen as a commissive speech act, defined as a speech act that the speaker intends to commit (Searle, 1975).[3] By contrast, when advising the buyer not to purchase insurance, the seller persuades the buyer to perform a specific action (e.g., not to take action to protect against potential losses). For Searle (1975), advice falls under the category of directive speech acts, defined as communication persuading another party to act in a particular manner. If a directive speech act is viewed as playing a more active role in the buyer's outcomes than a commissive speech act, omission bias and accountability theory predict that the advice mechanism can promote cooperation even when a bare promise is

[3] Reassuring messages that aren't promises, such as "I plan to invest", as seen in Bracht and Feltovitch (2009) would also be considered commissive speech acts, as the statement is about the speaker's intended action.

ineffective. Further, although a bare promise to ship the product may encourage more buyers to purchase the product and more sellers to cooperate with buyers, it is unclear how it affects insurance purchase decisions. One advantage of the insurance advice mechanism is that, in addition to promoting more transactions, it increases buyers' welfare by saving the cost of purchasing insurance—as observed in the comparisons between the IA treatment and the control. This could also be the difference between advising not to buy the insurance and other common advertising strategies used to persuade buyers to purchase a product. We discuss this in more detail at the end of the paper.

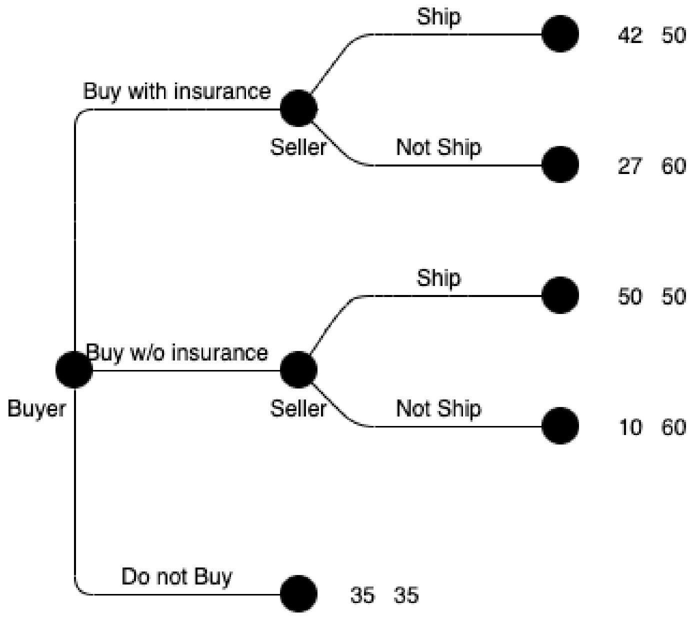
2. Experiment

2.1 Experimental design

Our experiment is based on a buyer-seller game (modified from Bolton et al., 2004; Li and Xiao, 2014). At the beginning of each treatment, subjects were randomly assigned to the role of either buyer or seller. Following Li and Xiao (2014), each treatment consisted of 10 rounds. Repeated games allowed us to obtain a larger number of observations and provided participants with opportunities to learn to converge to equilibrium. Both buyers and sellers received a full history of their decisions, which was provided and updated at the end of every round (see Appendix A for screenshots of the decision-making stage). To minimize the potential reputation effect, we randomly matched each buyer with a seller at the beginning of each round. At the end of the experiment, one round was randomly selected as the payment round, such that the earnings outcome in one round was unlikely to have any income effect on the decisions in later rounds. The instructions are provided in Appendix B.

In the control treatment (illustrated in Figure 1), at the beginning of each round, buyers and sellers were endowed with 35 points (experimental dollars), and the buyer could choose to purchase a product with insurance, purchase a product without insurance, or not purchase the product. The buyer valued the product at 40 points, which cost them 25 points to purchase. Following the previous literature (Li and Xiao, 2014; Lafky, 2014), we set the price of the product to be fixed to exclude the possibility that sellers could use the price to signal their intention to cooperate, which would complicate the study of the advice mechanism. If the buyer decided not to purchase the product, the round ended, and each participant's earnings remained at the 35 points endowment. If the buyer decided to purchase the product (with or without insurance), the seller received the payment of 25 points from the buyer and then decided whether to ship the product[4]. Shipping the product cost the seller 10 points. Thus, if the seller shipped the product, her earnings for that round were 50 points, and if she did not ship the product, her earnings were 60 points.

Figure 1: Buyer-seller game with insurance (control treatment)



The insurance cost the buyer 8 points. If the buyer purchased the product, but the seller did not ship it, the insurance would cover the loss of the 25 points that the buyer had paid to the seller. Once the buyer decided to purchase the insurance, he would pay the cost of 8 points, regardless of whether the seller shipped the product. All these factors were common knowledge.

The payoff structure was designed so that the buyer's decision (whether that be to purchase the product without insurance, purchase the product with insurance, or not purchase the product) differed depending on his belief in the likelihood that the seller would ship the product. This is discussed in more detail in Section 3.

The IA treatment—the timing of the game is described in Figure C1 in Appendix C—is the same as the control treatment, except that we added a stage before the buyer made their product and insurance purchase decisions. At this stage, the seller had to advise the buyer whether to purchase the insurance. The buyer then made his decision after he observed the

[4] Although we used the shipping context (also see Bolton et al., 2004; Li and Xiao, 2014), the nature of the decision making, however, can also extend to other settings such as the choice of the quality of the products or the speed of shipping.

seller's advice. The seller was informed of the buyer's insurance purchase decision before she made the shipping decision. All this was common knowledge. The rest of the game was the same as the control treatment.

The IA_HI treatment has the same structure as the IA treatment, except that the seller never knew whether the buyer decided to buy the insurance throughout the experiment.

2.2 Experimental procedure

The experiment was conducted at the Monash Laboratory for Experimental Economics (MonLEE) using z-tree (Fischbacher, 2007). The experimenter read the instructions aloud, and the subjects completed a comprehension quiz (see Appendix D) to ensure that they understood the task and the payoffs associated with each decision.

We ran 24 sessions in total—8 sessions per treatment—and we recruited, on average, 14 subjects in each session. Each session lasted less than one hour. Subjects were randomly assigned the role of either buyer or seller and maintained this role for the entirety of the experiment. In each round, a buyer was randomly and anonymously rematched with a seller. At the end of the experiment, one round was randomly selected as the payment round. In total, we recruited 332 subjects: 108 for the control treatment, 116 for the IA treatment, and 108 for the IA_HI treatment. Each subject was paid $4 AUD for participating, adding to the earnings from the games. The exchange rate was 1 point = $0.4 AUD. Subjects were paid privately, earning about $20 AUD on average.

3. Theoretical framework and hypotheses

In this section, we present a theoretical framework for deriving predictions for sellers' and buyers' decisions in each treatment. We start with a comparison between the control and the IA treatments. Later, we discuss the IA_HI treatment.

Consider a bilateral transaction between a buyer and a seller. The buyer demands one unit of the product that the seller produces and attaches a value ($v > 0$) to it. The seller attaches zero value to the product and can produce it at zero cost. We assume that both parties are risk-neutral. Following the previous literature (Lafky, 2014; Li and Xiao, 2014), the product's price is set as fixed in the experiment. Specifically, the product price is exogenously given by $p \in (0, v)$. The fixed price excludes the possibility that prices could be used as a signal for seller type and allows us to provide clean evidence for the effect of the insurance advice.[5]

If the buyer purchases the product, the seller can ship the product at a cost of $d \in (0, p)$. Following the standard approach of modelling seller reputation in an asymmetric information environment with both moral hazard and adverse selection problems (Bar-Isaac and Tadelis, 2008), we assume that there are two types of sellers: a good type (type-g) who always ships the product and a strategic type (type-s) who maximizes her own utility, including potentially a psychological cost, which we explain below. Only the seller knows her type. The buyer does not know the seller's type, but he does know that the probability of encountering a type-g seller is $q_g \in (0,1)$, and that the probability of encountering a type-s seller is $q_s = 1 - q_g$.

Along with purchasing the product, the buyer has the option to buy insurance at price w, which allows the buyer to recoup p in case the product is not shipped.[6] We assume that the insurance is not too expensive, that is, $w \leq p\left(1 - \frac{p}{v}\right)$, such that at least some buyers will buy the insurance in the control treatment.[7] The IA treatment has a special feature in that the seller can advise the buyer whether to buy the insurance before making any purchase decision. We denote the advice $a \in \{Y, N\}$, where Y means "buy the insurance" and N means "do not buy the insurance." Henceforth, we denote the seller's advice of not purchasing insurance as "*N*" and her advice of purchasing insurance as "*Y*." In contrast to the control treatment, in which the seller cannot influence the buyer's decisions, the seller's advice can change the buyer's expectation of the likelihood of receiving the product. As a result, the seller may become more accountable for the buyer's payoffs. Although the advice

[5] In the field, sellers may use price to signal their type and affect buyers' purchase decisions. In particular, high prices can be the efficient means of signaling high quality because a loss in sales could hurt low cost, low-quality sellers more than high cost, high-quality sellers (Bagwell and Riordan, 1991, and Bagwell, 1992). Assuming that a reasonable fraction of consumers can identify the true product quality before making purchases, if a low-quality seller pretends to be a high-quality seller by setting a high price, she will lose these well-informed buyers. That is, the loss in sales due to the high price will hurt low-cost, low-quality sellers more than high-cost, high-quality sellers, and high prices can then be used to effectively signal quality.

[6] We assume the seller does not receive commission from selling third-party insurances. If she does, the advice *N* should serve as an even stronger signal of being cooperative in the shipping stage. In reality, the buyer may also infer the distribution of the seller's type q_g from w. For example, the buyer may hold the belief of a low q_g in a market with high w. This will make him less likely to purchase the product in the control treatment, as we show below. However, it will not affect our hypotheses regarding the effects of the insurance advice mechanism.

[7] Without this assumption, buyers never buy insurance in the control. We keep this assumption to make sure insurance is not redundant. This assumption was satisfied in our experiment with w = 8, p = 25 and v = 40. In theory, if, instead, $w > p(1 - \frac{p}{v})$, buyers will not buy insurance in any of the three treatments because they receive a negative payoff. However, in this case, the advice mechanism is still beneficial in that it promotes more buyers to purchase the product and more sellers to ship the product. We discuss this in Appendix E4.

would not affect a type-g seller who always ships the product, we hypothesize a type-s seller will incur a psychological cost ($\alpha > 0$) for not shipping the product if (i) she advises N and (ii) the buyer purchases the product without insurance. The buyer does not know the exact value of α. However, for tractability, we assume he knows whether α is above or below d.

The seller may experience a psychological cost, even if the buyer purchases the insurance after she advises *N*. We assume the cost will be higher if the buyer follows the advice and does not purchase the insurance than purchase the insurance. In this sense, α can be understood as the incremental psychological cost between the two cases. Note that for simplicity, we also assume a type-s seller does not incur any psychological cost associated with the buyers' insurance purchase decisions (made without the sellers' influence) in the control treatment. That is, we assume that type-s seller's shipping decisions are not affected by buyers' insurance decisions in the control. This is because such a cost, if any, should be the same as in the IA treatment. This simplicity allows us to focus on the effect of the advice mechanism.

The timing of the game in the IA treatment is as follows. First, the seller advises whether to buy insurance. Next, the buyer receives the advice and decides whether to purchase the product and, if so, whether to buy the insurance. The seller observes the buyer's product purchase and insurance purchase decisions and decides whether to ship the product if the buyer purchases the product. The equilibrium concept is the weak perfect Bayesian equilibrium (WPBE). The formal propositions and proofs of our theoretical analysis can be found in Appendix E.

In the control treatment, without the advice stage, the buyer's purchase decision relies on the prior belief about the seller's type, q_g. The type-g seller always ships the product, whereas the type-s seller never ships the product. Given our assumption that $w \leq p\left(1 - \frac{p}{v}\right)$, it is straightforward to show that the buyer's optimal decision is as follows:

$$\begin{cases} \text{purchase the product without insurance,} & \text{if } q_g \geq 1 - \frac{w}{p} \\ \text{purchase the product with insurance,} & \text{if } \frac{w}{v-p} \leq q_g < 1 - \frac{w}{p} \\ \text{do not purchase the product,} & \text{if } q_g < \frac{w}{v-p} \end{cases} \quad (1)$$

That is, the buyer: purchases the product without insurance when q_g is relatively high; purchases the product with insurance if q_g is at some intermediate level; and does not purchase the product if q_g is very low.

Now consider the IA treatment. We note that the insurance advice mechanism is not effective when $d > \alpha$ as the type-s seller would rather incur the psychological cost than the shipping cost. The mechanism can only be effective with $d \leq \alpha$. The analysis below will focus on the case where $d \leq \alpha$.

It is straightforward to show that there is no separating equilibrium. If the type-s seller's separating-equilibrium advice is *Y*, the buyer's optimal choice is not to purchase the product, as he anticipates that the type-s seller will not ship the product. Thus, the type-s seller will advise *N* instead. If the type-s seller's separating-equilibrium advice is *N*, she will again be better off by instead advising *Y*, in which case she will get the full amount of payment p by not delivering, without incurring any psychological cost.

Next, we demonstrate that there exist two types of pooling equilibria when $d \leq \alpha$ in the IA treatment. First, there always exists a pooling equilibrium in which both types of sellers advise *N* (hereafter *N*-pooling equilibrium). Given that the type-g seller advises *N* in equilibrium, the type-s seller is better off pooling with them and advising *N*. As $d \leq \alpha$, the type-s seller's optimal choice after advising *N* is to ship the product and incur the shipping cost. Thus, buyers purchase the product without insurance upon receiving advice N, and both types of sellers advise *N* and subsequently deliver the product.

There may exist another pooling equilibrium with both types of sellers advising *Y* (hereafter *Y*-pooling equilibrium). The Y-pooling equilibrium exists only when $q_g \geq \frac{w}{v-p}$. When $q_g \geq \frac{w}{v-p}$, the type-s sellers do not want to deviate to advising *N* instead since advising *N* and revealing their type leads to a lower profit equal to $p - d$. The *Y*-pooling equilibrium does not exist if $q_g < \frac{w}{v-p}$. This is because buyers will not purchase the product after receiving advice *Y* and the type-s seller is better off deviating to advising *N*. Upon observing the off-path advice *N*, buyers will buy the product because they know that the seller will ship the product irrespective of her type given $d \leq \alpha$.

The above analysis suggests that both types of pooling equilibria may exist when $q_g \geq \frac{w}{v-p}$. The type-s seller prefers the *Y*-pooling equilibrium to the *N*-pooling equilibrium as in the *Y*-pooling equilibrium buyers still buy the product, but the type-s seller can choose not to ship without incurring any psychological cost. However, the *Y*-pooling equilibrium can be ruled out by applying forward induction. [8] The basic intuition is that, upon observing the off-path advice *N*, the buyer should believe that the seller is of type-g, as this seller is more likely to choose *N* than the type-s seller given only the latter incurs a psychological cost. Given that the *Y*-pooling equilibria may not always exist and can be ruled out by forward induction whenever it exists, we select the *N*-pooling equilibrium throughout the IA treatment.

Suppose the *N*-pooling equilibrium is selected. Compared to the control, the insurance advice mechanism always promotes efficient trades, as long as $d \leq \alpha$ for some sellers. Specifically, when $q_g < \frac{w}{v-p}$, trade takes place in IA while no trade takes place in the control. When $q_g \geq \frac{w}{v-p}$, buyers purchase the product in both the IA and control treatments. Both types of sellers deliver the products in IA while the type-s seller does not do so in the control. So the values from transactions are only realized in the IA but not in the control. In addition, buyers will not buy the insurance and thus save on the insurance cost.

Unlike the IA treatment, in the IA_HI treatment, the seller does not observe the buyer's insurance purchase decision. Intuitively, the equilibrium outcome remains the same as in the control when $d > \alpha$. The more interesting case arises when $d \leq \alpha$. First, recall that if the proportion of type-g sellers is sufficiently high such that $q_g \geq 1 - \frac{w}{p}$, buyers will buy the product without insurance in the control treatment. Thus, even though the seller is not informed of the buyer's insurance purchase decision, she anticipates that the buyer will not purchase the insurance. As a result, there exists an *N*-pooling equilibrium: both types of sellers advise *N*, all buyers purchase the product without insurance, and both types of sellers ship the product. The *Y*-pooling equilibrium can be again eliminated using the same forward induction argument as in the IA treatment.

Now consider the case when $q_g < 1 - \frac{w}{p}$. In this case, there can be multiple equilibria in the subgame after sellers advise *N*: (i) the buyer does not buy the product and the type-s seller does not ship the product upon receiving orders, (ii) the buyer buys the product without insurance and the type-s seller ships the product upon receiving orders, and (iii) the buyer mixes between buying the product with and without insurance, while the type-s seller mixes between shipping and not shipping the product. Specifically, in (iii), the buyer will buy the product without insurance with a probability of $\frac{d}{\alpha}$, while the type-s seller will ship the product with a probability of $1 - \frac{w}{p(1-q_g)}$. This mixed-strategy equilibrium exists given our

[8] The details of the equilibrium refinement using forward induction can be found in the Appendix E2.

assumption that the insurance price is relatively low, i.e., $w \leq p\left(1-\frac{p}{v}\right)$. We select the mixed-strategy equilibrium that, unlike other equilibria, does not require extreme forms of participants' (mis-)coordination when the insurance purchase decisions are unobservable.[9] In this mixed-strategy equilibrium, the proportion of each strategy realization is endogenously determined and can be used to make comparisons with other treatments using the data from the experiment. As in the IA treatment, we will focus on the equilibrium in which both types of sellers advise *N*. If all sellers advise *Y*, there will be no mixed-strategy equilibrium in the subsequent subgame as the type-s seller will never ship the product.

By comparing the equilibrium in all three treatments, we derive the following hypotheses. Note that our theoretical results do not qualitatively depend on the assumption that there exist type-g sellers who always ship the product. The insurance advice mechanism works by forcing type-s sellers to commit to shipping the product, which does not depend on the existence of type-g sellers. We discuss the details in Appendix E4.

Hypothesis 1: In both the IA and IA_HI treatments, sellers will advise *N*.

As long as α is sufficiently large ($\alpha \geq d$) for some sellers:

Hypothesis 2: Both the IA and IA_HI treatments increase the frequency of buyers purchasing the product compared to the control treatment:

freq(*BuyProd*|*Control*) <*freq* (*BuyProd* |*IA_HI*) = *freq*(*BuyProd*|*IA*).

Hypothesis 3: The frequency of buyers purchasing the product without insurance (*BuyProd & NoIns*) is highest in the IA treatment and lowest in the control treatment:

freq(*BuyProd&NoIns*|*Control*) < *freq*(*BuyProd&NoIns*|*IA_HI*) < *freq* (*BuyProd&NoIns*|*IA*).

Hypothesis 4: The frequency of sellers shipping the product is highest in the IA treatment and lowest in the control treatment:

freq(*Ship*|*Control*) < *freq*(*Ship*|*IA_HI*) < *freq*(*Ship*|*IA*).

[9] If sellers are optimistic and believe that buyers will follow the advice *N*, there exists an equilibrium such that when $d \leq \alpha$ all sellers advise *N*, buyers purchase the product without purchasing insurances, and all sellers ship the product. Similarly, if sellers are pessimistic and believe that buyers will not follow the advice *N*, there exists another equilibrium such that when $d \leq \alpha$ all participants behave as in the control treatment despite all sellers advising *N*. These equilibria are less convincing as they rely on sellers holding extreme beliefs (See Appendix E for more details).

These hypotheses imply that the insurance advice mechanism can improve market efficiency. We can measure market efficiency by the frequency of efficient trades. In our setup, there are two types of efficient trades. One is characterized by the number of buyers purchasing the product (regardless of their insurance purchase decisions) and the number of sellers shipping the product. The other is characterized by the number of buyers purchasing the product without insurance and the number of sellers shipping the product. When considering the total welfare of buyers and sellers, this latter definition of efficiency is a Pareto improvement over the first definition, as the buyer's earnings increase while the seller's earnings do not change. [10] To differentiate these two definitions of efficiency, we refer to the first type as *standard efficient trades* and to the second type as *optimal efficient trades*. According to Hypotheses 2, 3, and 4, we expect the frequency of efficient trades to be highest in the IA condition and lowest in the control condition. Consequently, the insurance advice mechanism is most effective in the IA condition, especially when considering the proportion of optimal efficient trades.

Hypothesis 5: The frequency of standard and optimal efficient trades *freq(Eff)*:
$freq(Eff|Control) < freq(Eff|IA\_HI) < freq(Eff|IA)$.

4. Results

We first report what advice sellers gave buyers in the two advice treatments. Then, we compare buyers' purchase decisions. Next, we examine how insurance advice affects sellers' shipping decisions and whether shipping frequency differs between the treatments. Lastly, we report the treatment effect on market efficiency. When reporting the findings, we always first report descriptive data followed by results from regression analysis to provide statistical tests for the comparisons. Since in each session we anonymously rematch participants after each round, the regression analysis allows us to control for the potential session effects, by clustering standard errors at the session level (Fréchette, 2012).

[10] Here we only consider the total welfare of buyers and sellers since the insurance is exogenously provided by the experimenter. In a natural setting, if the insurance market is close to perfectly competitive or insurers serve a large number of markets, a reduced number of insurance purchases in one market only leads to a negligible loss for the insurance company.

Table 1 summarizes the main descriptive results for each treatment. In total, there are 108 subjects in the control, 116 subjects in the IA treatment, and 108 subjects in the IA_HI treatment.

Table 1: Descriptive summary of decisions

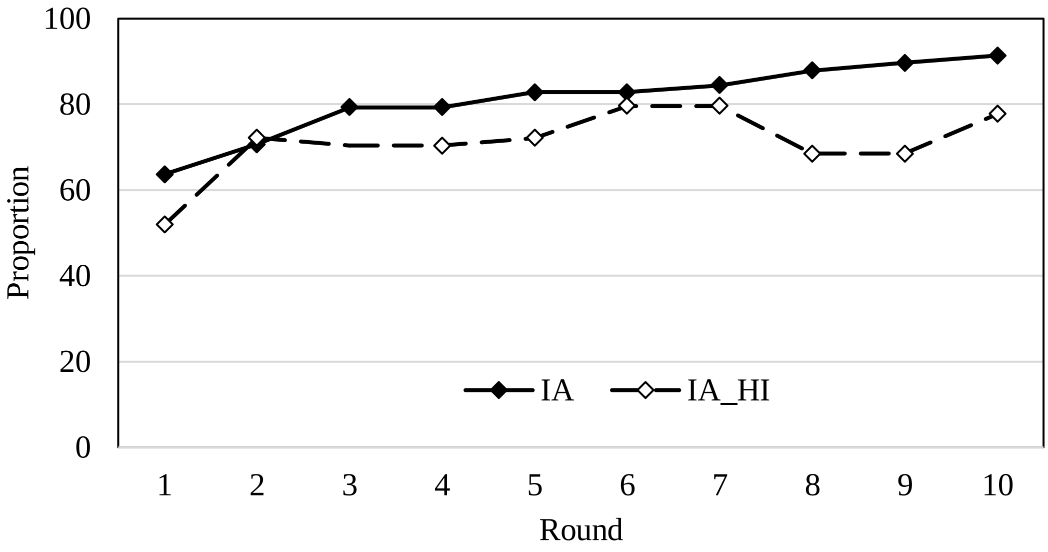
Treatment	Sellers advised N (%)	Buyers purchased product (%)	Buyers purchased product without insurance (%)	Sellers shipped product (%)	Standard efficient trades (%)	Optimal efficient trades (%)
Control	-	51.1	20.2	43.9	22.4	9.1
IA	81.2	66.9	48.3	61.4	42.1	33.3
Advice *N*	-	74.2	59.4	64.2	50.2	41.2
Advice *Y* *	-	35.0	2.7	25.0	9.5	0
IA_HI	71.1	67.8	35.6	53.5	37.8	22.8
Advice *N* *	-	73.2	43.6	58.0	46.2	29.4
Advice *Y**	-	52.0	13.6	26.2	16.7	5.4

Note: There are in total 54 buyers/sellers in the Control; 58 buyers/sellers in IA and 54 buyers/sellers in IA_HI. * The numbers of subjects in these three cases are smaller than the whole sample because some sellers never advised *N* or never advised *Y*, and some buyers never received advice *N* or never received advice *Y*. Specifically, in the IA treatment, 27 sellers never advised *Y* and 9 buyers who never received advice *Y*, leaving 49 buyers and 31 sellers in the Advice *Y* condition. In the IA_HI treatment, 4 sellers never advised *N* and 20 sellers never advised *Y*, leaving 54 buyers and 50 sellers in the Advice *N* condition, and 54 buyers and 34 sellers in the Advice *Y* condition.

4.1 Insurance advice

Supporting Hypothesis 1, over the 10 rounds, we observe a large proportion of sellers advised N in both the IA (81.2%) and the IA_HI treatment (71.1%). Figure 2 plots the proportion of sellers who advised *N* in each round. In both treatments, the frequency of advising *N* is relatively lower in the first round and increases over time.

Figure 2: Proportion of sellers who advised *N*



To provide statistical evidence for the treatment differences, we analyze sellers' insurance advice decisions over time using a random-effects linear probability.[11] We report the results in Table 2 below. The dependent variable is whether the seller advised *N* or *Y* in each round. Regressions (1) and (2) compare the IA treatment to the IA_HI treatment. The independent variable in Regression (1) includes only the IA treatment dummy variable. We find the coefficient of IA (β_1) is significantly positive, meaning that sellers in the IA treatment are significantly more likely to advise *N* than those in the IA_HI treatment. In Regression (2), we add the independent variables Round and IA*Round. We find that as the rounds progressed, sellers were significantly more likely to advise *N* in both the IA treatment ($\beta_2 + \beta_3$, p=0.001) and in the IA _HI treatment(β_2, p=0.032). We report in the next section that a buyer was more likely to purchase the product when he received advice *N* compared to when he received advice *Y*. The increasing rate of advising *N* suggests that sellers gained experience and learned to advise buyers *N* over time.

[11] For all the regressions reported in this paper, we cluster standard errors at the session level. We also test the robustness of the results by 1) including control variables that include gender, major and how well the individual understood the experiment instructions. All the main results are robust when including these controls (see Appendix F). We also conduct probit regressions, and all results are robust (see Appendix H).

Table 2: Random individual effects LPM regression analysis of insurance advice decisions

Independent variables	Dependent variable: Advice $N_{i,t}$ =1, if seller *i* advised *N* in round *t* *=0, o.w.*	
	(1)	(2)
β_1: IA	0.101**	0.032
	(0.040)	(0.084)
β_2: Round		0.014**
		(0.007)
β_3: IA_Round		0.013
		(0.011)
Constant	0.711***	0.633***
	(0.023)	(0.046)
H0: $\beta_2 + \beta_3$=0		p = 0.001
N	1,120	1,120

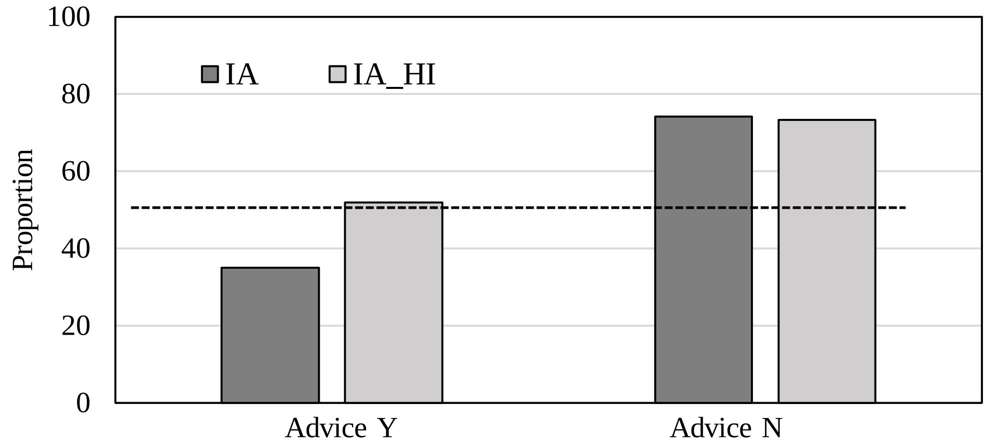
Note: IA_HI is the constant. Robust standard errors clustered at the session level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Result 1: *In both the IA and IA_HI treatments, the majority of sellers advised N. The frequency of advising N increases over time in both treatments.*

4.2 Purchase decision

Supporting Hypothesis 2, the purchase rate is higher in both the IA (66.9%) and IA_HI (67.8%) treatments compared to the control treatment (51.1%). Figure 3 plots the product purchase decision conditional on the advice that buyers received. As buyers did not receive any advice in the control treatment, we use the dotted line to mark the average product purchase rate in the control. Compared with the control, buyers in the two advice treatments were more likely to purchase the product when the sellers advised *N*. (IA: 74.2%; IA_HI: 73.3%; Control: 51.1%). By contrast, when buyers received advice *Y* in the IA treatment, the purchase rate was lower than in the control (IA: 35.0%; Control: 51.1%). In the IA_HI treatment, the purchase rate when advice is Y is about the same as that in the control (51.9%). These results suggest that the increase in the product purchase rate in the IA and IA_HI treatments is mainly driven by buyers who received advice *N*. When the seller advised *N*, the buyer's expectations of the seller shipping the product increased, and consequently, were more likely to purchase the product.[12]

Figure 3: Product purchase rate conditional on advice



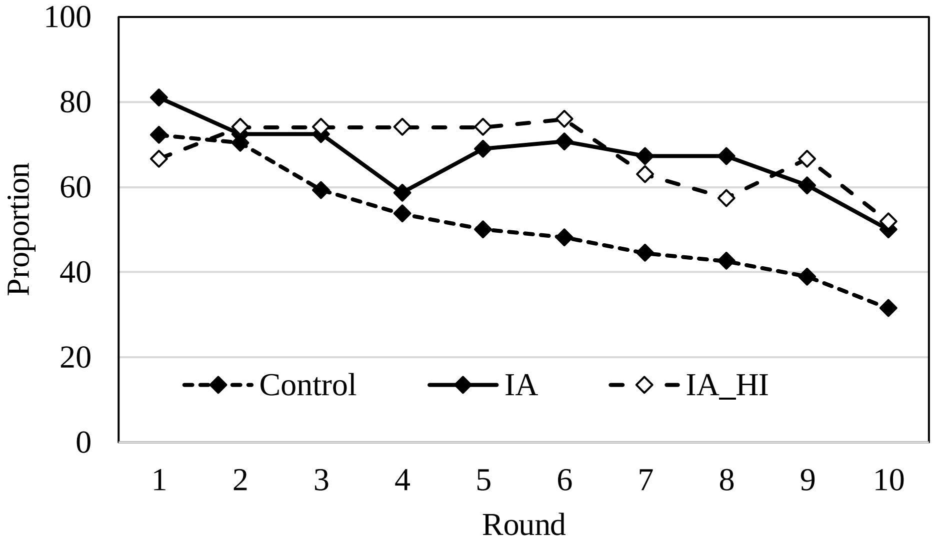
Note: The dotted line marks the purchase rate (51.1%) in the control treatment.

Next, we compare the dynamics of the product purchase decisions in each treatment. Figure 4(a) plots the proportion of product purchases over 10 rounds, while 4(b) and 4(c) plot the same proportion when the advice was *N* or *Y*, respectively. Although we observe a rapid decay in the purchase proportion in the control treatment, the decay is relatively slower in the IA and IA_HI treatments. Upon separating the cases based on sellers' insurance advice, we find that the decay is slower only when the sellers advised *N*.

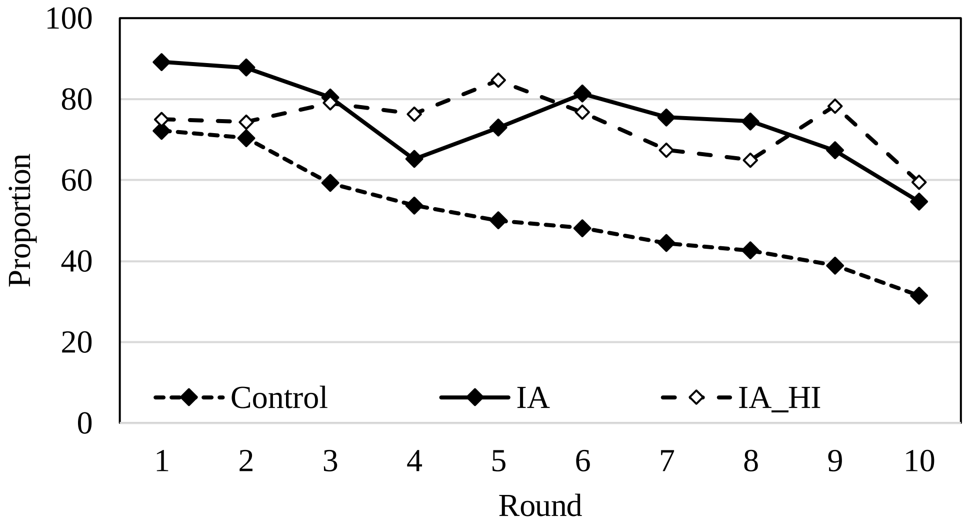
[12] It is interesting to observe that upon receiving advice *Y*, 35% of buyers in the IA and 51.9% of buyers in the IA_HI treatment purchased the product when in theory, they should not. However, the sample size is small; thus, caution must be observed when drawing inferences from these observations. Nevertheless, we make the following two notes. In the IA treatment, the proportion of buyers who purchased the product after receiving advice *Y* is relatively higher in the earlier rounds: 47.2% in the first five rounds and 13.5% in the last five rounds, with no buyers choosing to purchase the product in the final two rounds. This result suggests that buyers learned not to purchase the product when they received advice *Y* as the rounds progressed. In the IA_HI treatment, however, we still observe that 20% of buyers purchased the product in the last round. The three buyers who purchased the product in the IA_HI treatment when the seller advised *Y* appeared to be more cautious as all of them also decided to purchase insurance. By contrast, in the control treatment, 5 out of 17 buyers who bought the product in the last round did not purchase the insurance.

Figure 4. Proportion of buyers who purchased the product per round

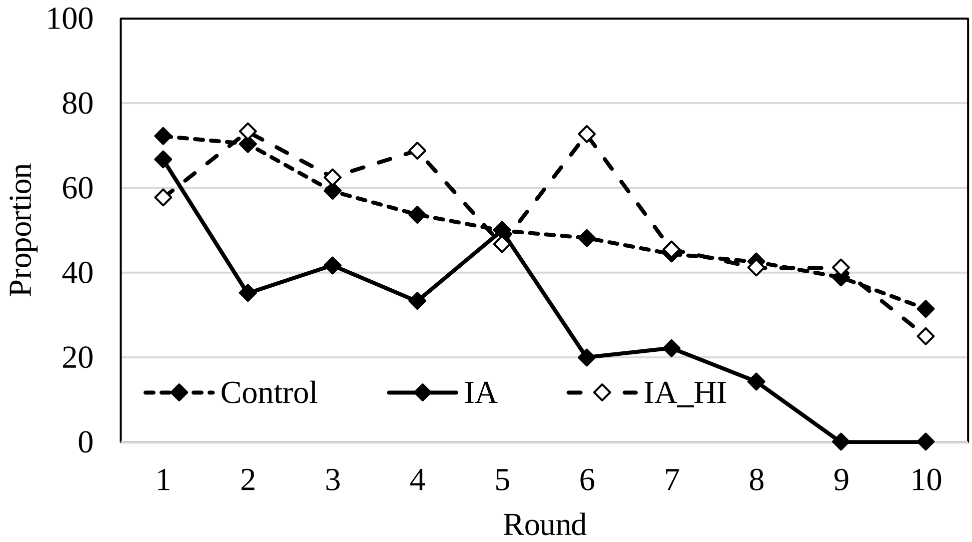
(a) Proportion of buyers who purchased the product per round (total)



(b) Proportion of buyers who purchased the product per round (Advice *N*)



(c) Proportion of buyers who purchased the product per round (Advice *Y*)



To provide statistical evidence for the comparisons reported above, we analyze the buyer's product purchase decisions (including when the advice was *Y*) using a random-effects linear probability model. We report the results in Table 3 below. The dependent variable is whether the buyer purchased the product in each round. Regressions (1)-(3) compare the IA treatment to the control. The independent variable in Regression (1) only includes the IA treatment dummy variable. We find the coefficient of IA (β_1) is significantly positive, meaning that buyers in the IA treatment are significantly more likely to purchase the product than those in the control. In Regression (2), the independent variables include the treatment dummy, round, and the interaction between the treatment and round. The coefficient of "Round" is negative and statistically significant, indicating that the product purchase rate decays significantly over time in the control treatment. There is also significant decay in the IA treatment ($\beta_3 + \beta_4$, $p = 0.026$). However, consistent with the observation in Figure 2, the IA treatment slows down the decay as compared to the control, the coefficient of the interaction variable IA*Round (β_4) is positive and marginally significant ($p = 0.093$). Regression (3) further includes the dummy variable Advice *N* to control for advice and the interaction variable Advice *N**Round. We find that the reduced decay rate in the IA treatment is mainly driven by sellers advising *N*. To see this, note that IA*Round is not statistically significant (and the direction is negative), indicating that advising Y did not slow down the rate of decay. By contrast, Advice *N* and the interaction variable Advice *N**Round are both positive and statistically significant.

Table 3: Random individual effects LPM regression analysis of product purchase decisions

Independent variables	Dependent variable: $Buy_{j,t}$ =1, if buyer *j* purchased the product in round *t* = 0, o.w.					
	(1) IA and Control	(2) IA and Control	(3) IA and Control	(4) IA_HI and Control	(5) IA_HI and Control	(6) IA_HI and Control
β_1: IA	0.158** (0.070)	0.045 (0.081)	-0.130 (.019)			
β_2: IA_HI				0.167* (0.089)	0.033 (0.089)	0.0111 (0.087)
β_3: Round		-0.042*** (0.007)	-0.042*** (0.007)		-0.042*** (0.007)	-0.042*** (0.007)
β_4: IA*Round		0.020* (0.012)	-0.018 (0.149)			
β_5: IA_HI*Round					0.024** (0.010)	-0.001 (0.014)
β_6: Advice *N*			0.289*** (0.079)			0.042 (0.058)
β_7: Advice *N* *Round			0.033*** (0.009)			0.033** (0.017)
Constant	0.511*** (0.065)	0.744*** (0.068)	0.744*** (0.068)	0.511*** (0.065)	0.744*** (0.068)	0.744*** (0.068)
H0: $\beta_3 + \beta_4$=0		p = 0.026	p < 0.001			
H0: $\beta_3 + \beta_5$=0					p = 0.012	p < 0.001
N	1120	1120	1120	1080	1080	1080

Note: Advice *N* = 1 if the seller advised *N*; = 0, o.w. Robust standard errors clustered at the session level are reported in the parentheses, *** p <0.01, ** p <0.05, * p <0.1.

Regressions (4)-(6) provide similar analyses of the IA_HI treatment compared to the control. The coefficient of IA_HI (β_2) is significantly positive in Regression (4), meaning that buyers in the IA_HI treatment are significantly more likely to purchase the product than those in the control. Regression (5) shows that although there is still a significant rate of decay in IA_HI ($\beta_3 + \beta_5$, p=0.012), it is slower in the IA_HI treatment compared to the control (β_5 is significantly positive). Similar to the findings of the IA treatment, when we control for the advice in Regression (6), we find that the reduction of the decay rate in the IA_HI treatment is significant only when sellers advised *N* (β_5 is negative and not significant; β_7 is significantly positive).

These results are consistent with our theoretical analysis, in which, for both the IA and IA_HI treatments, advice *N* leads to a higher product purchase rate compared to the control treatment.

We next tested Hypothesis 3 by comparing the proportion of buyers that purchased the product without insurance across treatments. Assuming that sellers shipped the product, buyers achieved the highest earnings in this scenario. Supporting Hypothesis 3, the proportion of buyers that purchased the product without insurance is highest in the IA treatment and lowest in the control treatment. The order is significant (IA: 48.3%; IA_HI: 35.6%; Control: 20.2%; Jonckheere–Terpstra test, $p < 0.001$).[13] We report the dynamics of this proportion over the 10 rounds in each treatment in Figure G1 in Appendix G. As shown in Figure G1, the order is very similar over the 10 rounds. This result suggests that one benefit of the IA mechanism is that buyers saved expenses on insurance (without increasing the risk of losing payment when encountering strategic sellers, as reported below).

Figure 5 further shows the proportion of buyers who purchased the product without insurance, conditional on the advice received. Since there was no advice opportunity in the control treatment, we mark the average proportion of product purchases without insurance by the dotted line. As shown in Figure 5, compared with the control treatment, the proportion of buyers who purchased the product without insurance after receiving advice *N* in both the IA and the IA_HI treatments more than doubled (IA: 59.4% IA_HI: 43.6%; Control: 20.2%). To test the treatment differences, we conducted a random-effects linear probability regression analysis of the insurance purchase decisions for IA and IA_HI, respectively. In each regression, we include only the treatment dummy variable (IA or IA_HI) and use the control as the baseline. In both regressions, the coefficient of the treatment dummy is significantly positive (IA: 0.39, $p=0.001$; IA_HI: 0.24, $p<0.001$). By contrast, when receiving advice *Y*, fewer buyers purchased the product without insurance in the IA treatment (3.7%) and the IA_HI treatment (13.6%) compared to the control treatment (20.2%,).[14] These results show that the lower insurance purchase rates in the IA and the IA_HI treatments reported above are due to buyers receiving advice *N*.

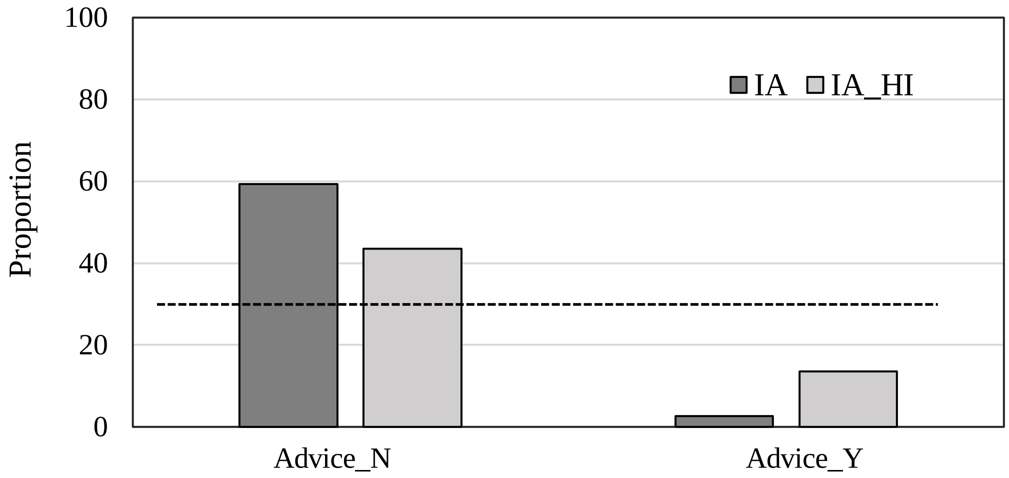
[13] For all Jonckheere-Terpstra tests, we calculate the average at the session level and use each session as an independent observation (n=8 for each treatment respectively, n=24 in total).

[14] We also tested the treatment differences using a similar regression analysis as above. We find the coefficient of the IA dummy is significantly negative (IA: -0.17, $p<0.001$) and the coefficient of the IA_HI dummy is negative, but not significant (IA_HI: -0.06, $p=0.218$).

Result 2: *Buyers were more likely to purchase the product in the IA and IA_HI treatments than in the control treatment. This increase in the product purchase rate is mainly driven by advice N.*

Result 3: *Buyers were more likely to purchase the product without insurance in the two advice treatments than in the control treatment. The increase is mainly driven by the effect of advising N.*

Figure 5: Proportion of buyers who purchased the product without insurance conditional on advice



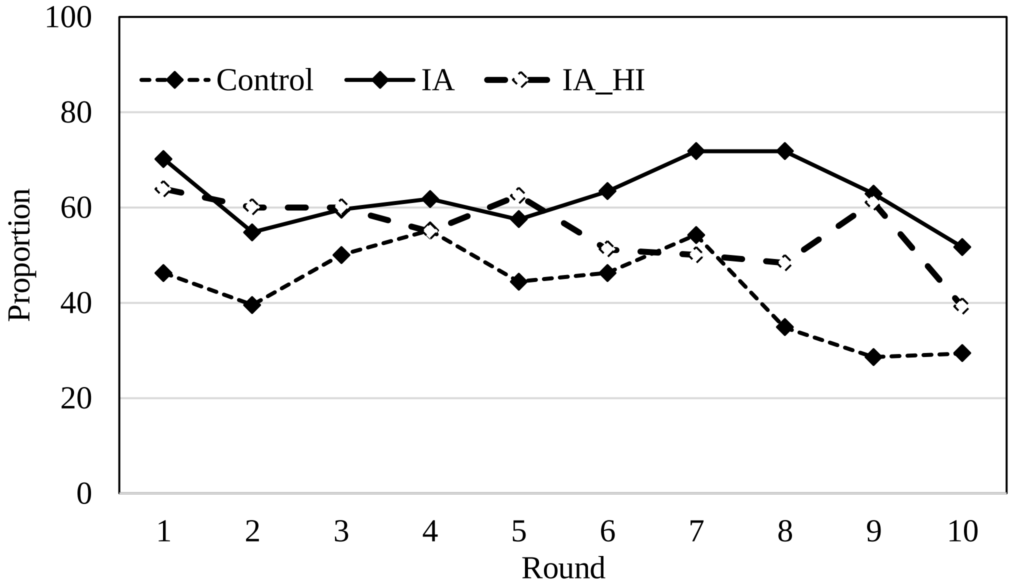
Note: The dotted line marks the product purchase rate (20.2%) in the control treatment (no advice was given in the control).

4.3 Shipping decision

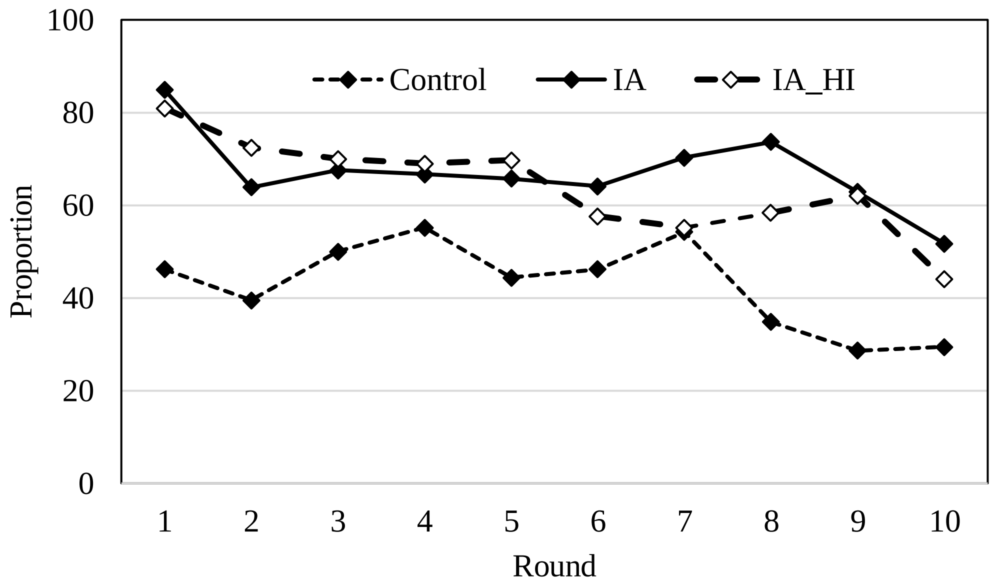
Supporting Hypothesis 4, the overall shipping rate is highest in the IA and lowest in the control treatment (IA: 61.4%; IA_HI: 53.5%; Control: 43.8%, Jonckheere–Terpstra test, p= 0.006). The increase in the shipping rate is mainly driven by advising *N* (IA (advice N): 64.2%; IA_HI (advice N): 58.0%; Control: 43.9%). By contrast, when sellers advised *Y*, the shipping rate in the two advice treatments is lower than in the control (IA (advice Y): 25.0%; IA_HI (advice Y): 26.2%; Control: 43.9%). Figure 6 plots the shipping rate over the 10 rounds. Figure 6(a) reports the overall shipping rate. We observe that the overall shipping rate is highest in the IA treatment in 7 out of the 10 rounds. By separating the cases into advice *N* (Figure 6(b)) and advice *Y* (Figure 6(c)), we again find that the higher shipping rate in the IA treatment is mainly driven by advising *N* throughout the experiment. These results are consistent with our theoretical framework, in which the seller incurs a psychological cost for not shipping the product after she advised *N*.

Figure 6. Proportion of sellers who shipped the product in each round

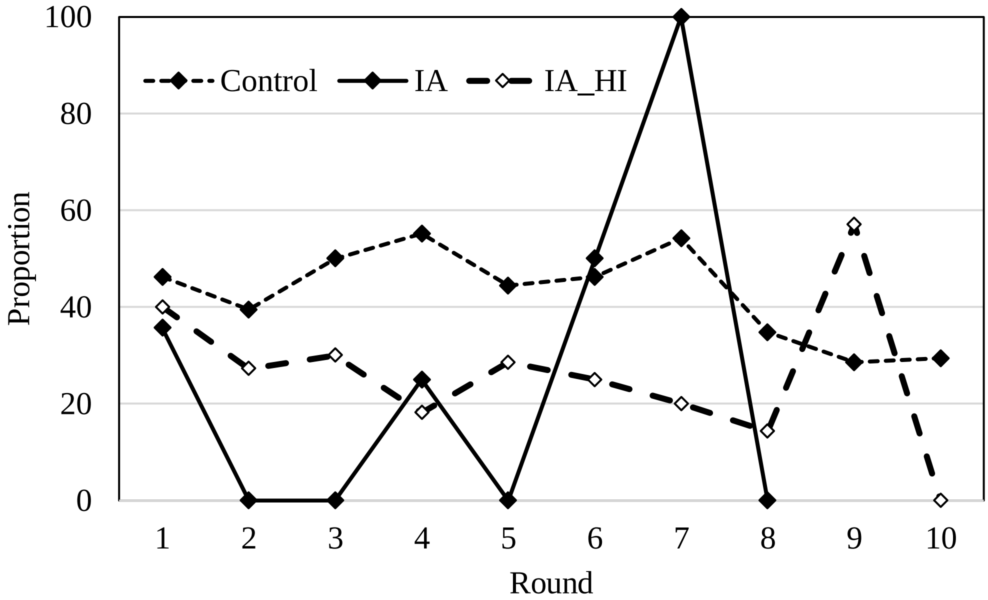
(a) Proportion of sellers who shipped the product in each round (Total)



(b) Proportion of sellers who shipped the product in each round (Advice *N*)



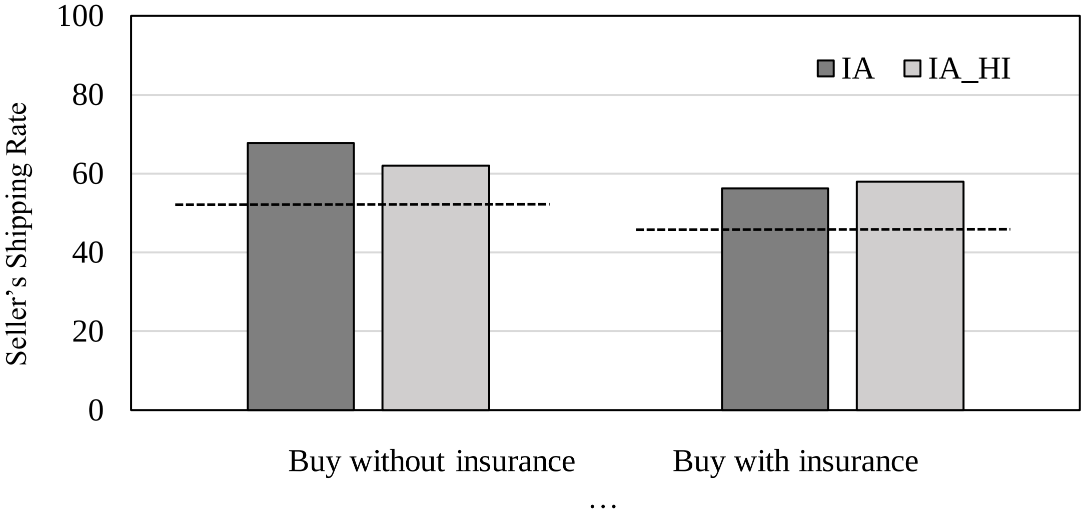
(c) Proportion of sellers who shipped the product in each round (Advice *Y*)



Note: In rounds 9 and 10 of the IA treatment, there was no observation in which the seller advised *Y* and the buyer purchased the product.

In our theoretical framework, we assume that the psychological cost is higher when the seller observes that the buyer follows her advice of *N* than when the buyer does not follow her advice. This assumption would predict that, in the IA treatment, sellers who advise *N* would be more likely to ship the product after observing the buyer who purchased the product without insurance than buyers who purchased the product with insurance. To check this, Figure 7 reports the average shipping rates of sellers who advised *N* in the IA treatment when buyers purchased the product with and without insurance. For comparison, the dotted lines mark the average shipping rate when buyers purchased the product with and without insurance in the control treatment. We also include data from the IA_HI treatment. As sellers never knew the buyer's insurance purchase decision in the IA_HI treatment, we do not expect to see any correlation between the sellers' shipping decisions and the buyers' insurance purchase decision. Our data are consistent with the prediction. The average shipping rate in the IA treatment is higher when the buyer did not purchase the insurance (67.8%) than when the buyer purchased insurance (56.2%). The difference is smaller in the IA_HI treatment and in the control.

Figure 7: Shipping rates when sellers advised *N*



Note: The dotted line marks the shipping rate in the control treatment when the buyer did not purchase an insurance (48.4%, on the left) and when the buyer purchased insurance (42.0%, on the right).

To provide statistical evidence for the above comparisons, we conduct a regression analysis of the sellers' shipping decisions using a random effects linear probability model. For each of the advice treatments, we start with a regression that only includes the treatment dummy (IA or IA_HI) (Regressions 1 and 4). Results are reported in Table 4. the results from Regression (1) show that, in the IA treatment, sellers are significantly more likely to ship than those in the control (β_1 is statistically significant). However, in Regression (4) while we observe the direction is higher in IA_HI as compared to the control: β_2 is positive but not significant.[15] Next, we add another independent variable "Advice *N*" = 1 if the seller advised *N* (Regression 2 and 5). Results from the two regressions show that the insurance advice mechanism of advising N is what drives the increase in shipping rates (β_6 is statistically significant in both treatments).

To test the effect of buyers' insurance purchase decisions, in Regressions (3) and (6) we add the independent variable "Noinsure"=1 if the buyer purchased the product without insurance and the interaction of the Noinsure variable with the dummy treatment variable. As our hypothesis on the psychological cost is only for sellers who advised *N*, we do not include

[15] We also ran a similar regression model using only data from IA and IA_HI with IA_HI being the baseline, we find that the coefficient of the IA dummy is positive but not significant (IA: 0.08, p=0.232)

data from the advice treatments when the advice was *Y*.[16] Regression (6) shows that the shipping rate is not affected by the insurance purchase decision in the control treatment (β_3 is not statistically significant), or in the IA_HI treatment ($\beta_3 + \beta_4$, p=0.204). In contrast, results from Regression (3) suggest that, in the IA treatment, if the buyer purchased the product without insurance after receiving advice N, sellers are significantly more likely to ship the product ($\beta_3 + \beta_4$, p<0.001).

Table 4: Random individual effects LPM regression analysis of shipping decisions

Independent variables	Dependent variable: $Ship_{i,t}$ =1, if the seller *i* shipped the product in round *t*; =0， o.w.					
	(1) IA and Control	(2) IA and Control	(3) IA (advice *N*) and Control	(4) IA_HI and Control	(5) IA_HI and Control	(6) IA_HI (advice *N*) and Control
β_1: IA	0.179*** (0.053)	-0.062 (0.073)	0.096 (0.066)			
β_2: IA_HI				0.101 (0.0063)	-0.059 (0.073)	0.159** (0.078)
β_3: Noinsure			0.046 (0.058)			0.046 (0.058)
β_4: Noinsure*IA			0.121* (0.064)			
β_5: Noinsure*IA_HI						-0.018 (0.063)
β_6: Advice *N*		0.273*** (0.061)			0.216*** (0.074)	
Constant	0.436*** (0.036)	0.436*** (0.036)	0.419*** (0.049)	0.437*** (0.036)	0.437*** (0.049)	0.419*** (0.036)
H0: $\beta_3 + \beta_4 = 0$			p < 0.001			
H0: $\beta_3 + \beta_5 = 0$						p = 0.204
N	664	664	625	642	642	558

Note: Noinsure = 1 if the buyer did not purchase insurance; = 0, o.w. Advice *N* = 1 if the seller advised *N*; = 0, o.w. Robust standard errors clustered at the session level are reported in parentheses. *** p <0.01, ** p <0.05, * p <0.1.

In summary, our data suggest that advice *N* provided by the sellers causes an increase in the shipping rate. Furthermore, for sellers, the buyers’ insurance purchase decisions did

[16] Similar regression analysis using only data from the advice treatments when the advice was *Y* shows that in the IA treatment sellers were significantly less likely to ship the product when buyers did not follow the advice *Y* and did not purchase insurance than the control ($\beta_3 + \beta_4 = -0.146$, p=0.028). There is no significant difference in shipping rates between the control and the IA_HI treatment when the buyer purchased the product without insurance ($\beta_3 + \beta_4 = 0.100$, p=0.573).

matter when they advised buyers *N*, as long as they could observe whether the advice was followed.

We also explore individual differences in shipping behavior. For each seller, we calculate the frequency of shipping the product when the paired buyer decided to purchase the product. The distribution of the shipping rate in each treatment is shown in Appendix G (Figures G2). In all treatments, the two most common behavior profiles are to never or always ship. Table 5 summarizes the proportion of sellers who "always shipped" and "never shipped" in each treatment. In the two advice treatments, we report the proportions for the case when sellers advised *N* and *Y*, respectively.[17]

Our theoretical framework predicts that a type-s seller who never ship in the control treatment will advise *N* and subsequently always ship in the IA treatment, provided that the psychological cost of not shipping is sufficiently large. In the IA_HI treatment, there is a mixed strategy equilibrium where type-s sellers advise *N*, but only some will always ship the product. Thus, compared to the control treatment, we expect to see a higher proportion of "always ship" and a lower proportion of "never ship" when sellers advise *N* in the IA treatment. The effect of advice, albeit positive, is weaker in the IA_HI treatment. We thus compare the proportion of "always ship" and "never ship" when sellers advise *N* in the two advice treatments with the control treatment.

Supporting the theoretical framework, we find that in IA and IA_HI, when the advice was *N*, the proportion of "always ship" is highest in the IA treatment and lowest in the control, and the order is statistically significant (41.4% >34.0% >18.5%, Jonckheere–Terpstra test, $p = 0.003$). Similarly, the proportion of "never ship" is highest in the control and lowest in the IA treatment, and the order is also statistically significant (33.3 >27.7% >19.0%, Jonckheere–Terpstra test, $p = 0.033$).[18] [19]

[17] Calculating the two types using the pooled data in the two advice treatments can be misleading. Take an extreme case as an example. Suppose all sellers advise *N* in 8 rounds and advise *Y* in 2 rounds. Also suppose all sellers always ship when advising *N* and never ship when advising *Y*. In this case, when calculating the overall proportion of "always ship" and "never ship", we will get 0% of both "always ship" and "never ship" in the advice treatments.

[18] Sellers behaved very differently when they took the off-equilibrium strategy of advising *Y*. The proportion of "always ship" is highest in the control treatment and lowest in the IA_HI (Control: 18.5%; IA: 10%; IA_HI: 4%). For the proportion of "never ship", it is highest in the IA and lowest in the control (IA: 60%; IA_HI: 48.6%; Control: 33.3%)

[19] In the IA treatment, we find that 11 out of 58 sellers never shipped when advising N. Among these 11 sellers, 5 only advised N. Among the other 6 sellers who sometimes advised N and sometimes advised Y, 5 of them never shipped and the other one shipped 50% of the time after advising Y. Thus, we may argue at most, 10 out of 58 (17%) sellers had $d > \alpha$. This indicates that there is a small proportion of sellers who had $d > \alpha$.

Result 4: *Sellers were more likely to ship the product in the two advice treatments than in the control. The increase in the shipping rate is driven by those who advised N.*

Result 5: *In the IA treatment, sellers who advised N were more likely to ship the product when the buyer followed the advice than when he did not follow the advice.*

Table 5: Frequency of sellers who either always or never shipped the product

Treatment (# of sellers)	Always Ship (%)	vs. Control (p-value)	Never Ship (%)	vs. Control (p-value)
Control (54)	18.5	-	33.3	-
IA (58)	32.8	0.062	17.2	0.043
Advice_*N* (58)	41.4	0.004	19.0	0.082
Advice_*Y* (20)	10.0	0.426	60.0	0.042
IA_HI (54)	24.0	0.468	24.0	0.303
Advice_*N* (47)	34.0	0.057	27.7	0.574
Advice_*Y* (25)	4.0	0.002	48.0	0.176

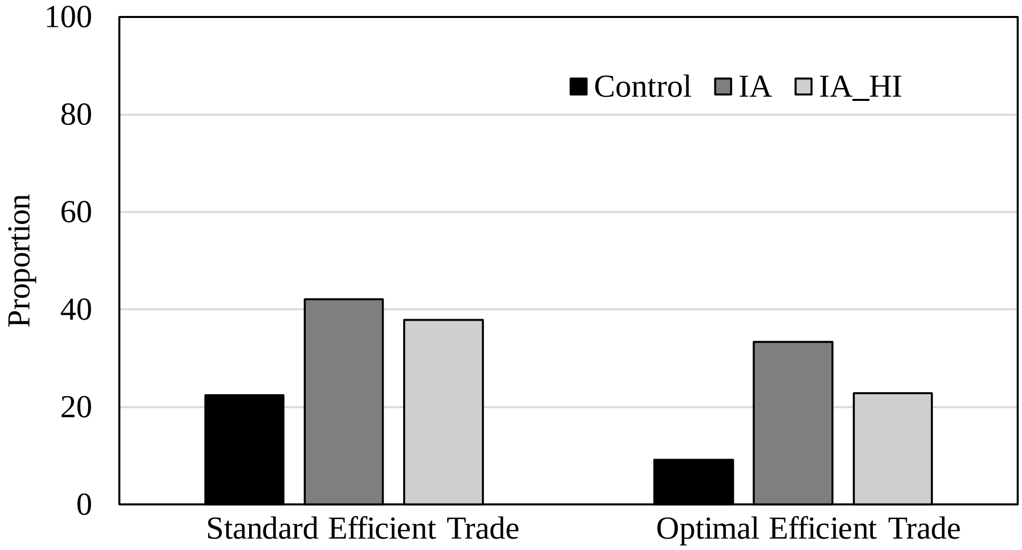
Note: For Advice_*N* (Advice_*Y*), we only considered the seller's shipping decisions when the advice was *N* (*Y*) and a buyer purchased the product. The p-value was based on an LPM regression analysis with standard errors clustered at the session level, where the independent variable is either a treatment dummy, or an advice dummy[20].

4.4 Market Efficiency

Figure 8 plots the proportion of standard efficient trades and optimal efficient trades in each treatment. As defined in Section 3, standard efficient trade occurs when the product is bought and shipped, and optimal efficient trade occurs when the product is bought without insurance and shipped. Figure 8 supports Hypothesis 5 in that the frequency of standard efficient trades is highest in the IA and lowest in the control treatment (IA: 42.1%; IA_HI: 37.8%; Control: 22.4%, Jonckheere–Terpstra test, $p= 0.023$). and we observe the same pattern for optimal efficient trades (IA: 33.3%; IA_HI: 22.8%; Control: 9.1%, Jonckheere–Terpstra test, $p<0.001$).

[20] When comparing between IA and IA_HI using similar regression analysis with IA_HI being the baseline, we find the sign of the coefficient of the IA dummy treatment is consistent with our hypothesis, although it is not significant: Always ship (IA: 0.07, p=0.338; when include only advicd N: IA: 0.09, p=0.217); Never ship (IA: -0.09, p=0.326; when include only advice *N:* IA: -0.07, p=0.378).

Figure 8: Proportion of efficient trades by treatments



We also conduct a random-effects linear probability regression analysis of efficient trades to test the effect of IA and IA_HI, respectively. In each regression, we include only the treatment dummy variable (IA or IA_HI) and use the control as the baseline. We find the coefficient for each treatment dummy is significantly positive (standard efficient trades: IA: 0.20, p=0.001; IA_HA: 0.15, p=0.030; optimal efficient trades: IA: 0.24, p<0.001; 0.14, p=0.013), indicating that the increase in the efficient trades in both treatments is statistically significant. In addition, we compare the efficient trades between IA and IA_HI treatments using a similar regression analysis where we include only the IA dummy and use IA_HI as the baseline. We find the coefficient of IA is positive for both standard efficient trades (0.04, p=0.531) and optimal efficient trades (0.10, p=0.093) although only the latter is marginally significant.

As a result of the increase in the number of efficient trades, both buyers and sellers made higher profits in the two insurance advice treatments than in the control. The average earnings per round for the buyer increase by 9.1% in the IA treatment (36.1 points) compared to the control (33.1 points). The average earnings per round for the seller increase by 4.4% in the IA treatment (47.5 points) compared to the control (45.5 points). We observe similar results in the IA_HI treatment. Compared to the control, the average earnings per round for the buyers increase by 5.4% (34.9 points), and the average earnings per round for the seller increase by 5.9% (48.2) in the IA_HI treatment.

Result 6: *The insurance advice mechanism significantly increases the frequency of efficient trades, especially when buyers' insurance purchase decisions were known to the sellers.*

5. Discussion and Conclusion

We designed and tested a novel insurance advice mechanism aimed at promoting efficient trade in a market with asymmetric information. We show both theoretically and experimentally that under this mechanism, buyers purchased the product significantly more often, and sellers were also more likely to ship the product than in the control treatment. The insurance advice mechanism also has an indirect welfare effect on buyers by reducing the frequency of purchasing insurance. The comparison between the two advice treatments further suggests that the mechanism is most effective when sellers could observe buyers' insurance purchase decisions. This finding suggests that online marketplaces may want to make buyers' insurance purchase decisions salient to sellers, alongside the introduction of the insurance advice mechanism.

Our study points out a new direction for designing market mechanisms to overcome asymmetric information problems. Existing instruments, such as reputation mechanisms, warranties, and insurance, designed to promote market efficiency, are often costly to sellers and/or buyers (Li and Xiao, 2014; Bolton et al., 2018; Bolton et al., 2019; Andreoni, 2018). Although some big companies or manufacturers can provide insurance or warranties, many small sellers that populate online marketplaces, such as eBay and Amazon, cannot do so. Sellers can sometimes use prices to signal their quality. However, this type of signaling can be quite costly: good sellers have to distort their prices to differentiate themselves from bad sellers (Bagwell and Riordan, 1991, Bagwell, 1992). Importantly, price signaling does not change the non-cooperative behavior of strategic sellers.

The insurance advice mechanism can be a low-cost complement to these existing instruments. A simple message on the seller's online page can do the job. With the rapid growth of the digital economy, more third-party insurance companies—such as Squaretrade and xcover. com—have emerged to protect consumers against risks not covered by the manufacturer's warranty. These insurance products provide a natural opportunity to introduce the advice mechanism with minimal changes to the current platform design. It is worth noting that in this paper, the benefits of the advice mechanism are measured using the outcomes in a market that already has third-party insurance as the reference point. Our experiment cannot speak for whether a market without third-party insurance can benefit from introducing "insurance" and "advice" mechanisms simultaneously. A mechanism that requires adding both insurance and advice is surely not as simple.

One may argue that even in a market with insurance, it might still be sufficient to simply advise the buyers to purchase the product. We make two remarks on this: First, the fact that sellers are selling their products online or in stores is probably equivalent to advising buyers to purchase the product. That is, the advice to buy the product is probably already embedded in the market and does not provide any new information (e.g., the quality of the product as argued in our paper). Advising whether consumers should purchase the insurance from a third party, in contrast, provides new information.

Second, in our framework, we assume that under the advice mechanism, sellers' psychological cost of not shipping depends on whether buyers follow the advice not to buy insurance. Consistent with this assumption, data from our experiment show that one advantage of the insurance advice mechanism is that, in addition to promoting more transactions, it increases buyers' welfare by saving the cost of purchasing insurance—as observed in the comparisons between the IA treatment and the control. If the seller's advice is simply to buy the product, buyers can follow the advice by purchasing the product with or without the insurance. Thus, we expect that advice to purchase the product will be less likely to help buyers save the cost of buying insurance. Future studies could test these hypotheses and compare the effects of the advice mechanism with other forms of communication, such as bare promises (both structured and free forms) or direct advice to buy products.

A potential barrier to implementing the advice mechanism is whether third-party insurance companies will agree to it. One concern may be that third-party insurance companies will be worse off if fewer people purchase insurance. However, it is unclear whether the introduction of the insurance mechanism necessarily makes the insurance company worse off. In the Advice_HI treatment, on average, more products were purchased with insurance (32.2%) than in the control treatment (30.1%). This means that if there is no observability, insurance advice may result in companies like Squaretrade being better off (or at least not worse off) because more buyers participate in the market. Finally, it is the online platform that would implement the mechanism, not the third-party insurance company. Assuming that online platforms such as eBay benefit from more buyers purchasing products, we should expect that there is a potential incentive to introduce the advice mechanism.

Future research could be valuable by examining factors that may impact the mechanism's effectiveness. A critical condition for the insurance advice mechanism to be effective is that the psychological cost of not shipping after advising no insurance is greater than the cost of shipping. In our experiment, the seller's shipping cost was relatively small.[21] It would be interesting to explore whether the psychological cost is sufficient to discourage non-cooperative behavior in markets where the cost of cooperation is much greater (e.g. when products are expensive). Previous studies on lying aversion have shown that although people are more likely to lie when they have more to gain, they are also less inclined to lie when the other person has more to lose (Gneezy, 2005; Lundquist et al., 2009). If the psychological costs increase with these gains from non-delivery, the advice mechanism may remain effective, even for relatively large-value transactions.

In our experiment, there was no uncertainty associated with receiving compensation under the insurance policy. However, asymmetric information problems are also common in the insurance market. In particular, consumers are often unsure about the coverage. It would be valuable to examine the relationship between the effect of the advice mechanism and trust in the insurance policy. Although we forced the sellers to provide advice in our experiment, it may be more feasible to introduce a mechanism in which insurance advice is presented as an option or upon request from the buyers and sellers can choose to be silent. It would be fruitful to compare the effect of the advice mechanism when advising is mandatory as opposed to when it is optional.

[21] Although the cost of shipping is only 10 points ($4 AUD), previous studies have shown that people are willing to incur a high cost to cooperation. For example, Abeler et al. (2019) find that people are willing to act honestly at a cost as high as $50 USD,

References

Abeler, J., Becker, A., and Falk, A. (2014). Representative evidence on lying costs. *Journal of Public Economics*, *113*, 96-104

Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, *87*(4), 1115-1153

Andreoni, J. (2018). Satisfaction guaranteed: When moral hazard meets moral preferences. *American Economic Journal: Microeconomics*, *10*(4), 159-89.

Balafoutas, L., and Sutter, M. (2017). On the nature of guilt aversion: Insights from a new methodology in the dictator game. *Journal of Behavioral and Experimental Finance*, *13*, 9-15.

Bagwell, K. (1992). Pricing to signal product line quality. *Journal of Economics and Management Strategy*, 1(1), 151-174.

Bagwell, K., and Riordan, M. H. (1991). High and declining prices signal product quality. *American Economic Review*, 224-239.

Bar-Isaac, H. and Tadelis, S. (2008). Seller reputation. *Foundations and Trends in Microeconomics*, *4*(4), 273-351.

Battigalli, P., Charness, G., and Dufwenberg, M. (2013). Deception: The role of guilt. *Journal of Economic Behavior and Organization*, *93*, 227-232

Battigalli, P., and Dufwenberg, M. (2007). Guilt in games. *American Economic Review*, *97*(2), 170-176.

Belot, Michèle, V. Bhaskar, and Jeroen Van De Ven. "Can observers predict trustworthiness?." *Review of Economics and Statistics* 94, no. 1 (2012): 246-259.

Bénabou, R., & Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, *30*(3), 141-164

Bicchieri, C., and Lev-On, A. (2007). Computer-mediated communication and cooperation in social dilemmas: an experimental analysis. *Politics, Philosophy and Economics*, *6*(2), 139-168

Binmore, K. (2006). Why do people cooperate? *Politics, Philosophy and Economics*, *5*(1), 81-96.

Bolton, G., Greiner, B., and Ockenfels, A. (2018). Dispute resolution or escalation? The strategic gaming of feedback withdrawal options in online markets. *Management Science*, *64*(9), 4009-4031.

Bolton, G. E., Katok, E., and Ockenfels, A. (2004). How effective are electronic reputation mechanisms? An experimental investigation. *Management Science*, *50*(11), 1587-1602.

Bolton, G. E., Kusterer, D. J., and Mans, J. (2019). Inflated reputations: Uncertainty, leniency, and moral wiggle room in trader feedback systems. *Management Science*, *65*(11), 5371-5391.

Bracht, J., and Feltovich, N. (2009). Whatever you say, your reputation precedes you: Observation and cheap talk in the trust game. *Journal of public economics*, *93*(9-10), 1036-1044

Brandts, J., Cooper, D. J., and Rott, C. (2019). Communication in laboratory experiments. In *Handbook of research methods and applications in experimental economics*. Edward Elgar Publishing.

Cabral, L., and Hortacsu, A. (2010). The dynamics of seller reputation: Evidence from eBay. *The Journal of Industrial Economics*, *58*(1), 54-78.

Cartwright, E. (2019). A survey of belief-based guilt aversion in trust and dictator games. *Journal of Economic Behavior and Organization*, *167*, 430-444.

Charness, G., and Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, *74*(6), 1579-1601.

Charness, G., and Dufwenberg, M. (2010). Bare promises: An experiment. *Economics Letters*, *107*(2), 281-283.

Chen, Y., Cramton, P., List, J. A., and Ockenfels, A. (2021). *Market Design, Human Behavior, and Management*. Management Science. 67(9). 5317-5348

Chen, Y., and Zhang, Y. (2021). Do elicited promises affect people's trust?—Observations in the trust game experiment. *Journal of Behavioral and Experimental Economics*, *93*, 101726

Cressey, D. R. (1986). Why managers commit fraud. *Australian and New Zealand Journal of Criminology*, *19*(4), 195-209

Dellarocas, C., and Wood, C. A. (2008). The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Science*, *54*(3), 460-476

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 67-80

Ellingsen, T., and Johannesson, M. (2004). Promises, threats and fairness. *The Economic Journal*, *114*(495), 397-420.

Erat, S., and Gneezy, U. (2012). White lies. *Management Science*, *58*(4), 723-733.

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

Fréchette, G.R. (2012) Session-effects in the laboratory. *Experimental Economics* 15, 1-14.

Gino, F., Norton, M. I., & Weber, R. A. (2016). Motivated Bayesians: Feeling moral while acting egoistically. *Journal of Economic Perspectives*, *30*(3), 189-212

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, *95*(1), 384-394

Gneezy, U., Rockenbach, B., and Serra-Garcia, M. (2013). Measuring lying aversion. *Journal of Economic Behavior and Organization*, *93*, 293-300.

Lafky, J. (2014). Why do people rate? Theory and evidence on online ratings. *Games and Economic Behavior*, *87*, 554-570

Li, L., and Xiao, E. (2014). Money talks: Rebate mechanisms in reputation system design. *Management Science*, *60*(8), 2054-2072.

Lerner, J. S., and Tetlock, P. E. (1994). Accountability and social cognition. *Encyclopedia of Human Behavior*, *1*, 3098-3121.

Lerner, J. S., and Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*(2), 255.

López-Pérez, R., and Spiegelman, E. (2013). Why do people tell the truth? Experimental evidence for pure lie aversion. *Experimental Economics*, *16*(3), 233-247.

Lundquist, Tobias, Tore Ellingsen, Erik Gribbe, and Magnus Johannesson. "The aversion to lying." *Journal of Economic Behavior and Organization* 70, no. 1-2 (2009): 81-92

Mayzlin, D., Dover, Y., and Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, *104*(8), 2421-55.

Resnick, P., and Zeckhauser, R. (2002). Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system. *The Economics of the Internet and E-commerce*, *11*(2), 23-25.

Ritov, I., and Baron, J. (1992). Status-quo and omission biases. *Journal of Risk and Uncertainty*, *5*(1), 49-61.

Sally, D. (1995). Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and society*, *7*(1), 58-92.

Sánchez-Pagés, S., and Vorsatz, M. (2007). An experimental study of truth-telling in a sender–receiver game. *Games and Economic Behavior*, *61*(1), 86-112.

Searle, J. R. (1975). ‘A taxonomy of illocutionary acts,’ in K. Gunderson (ed.), *Language, Mind and Knowledge*, Minneapolis, MN: University of Minnesota Press, 344–369.

Serra-Garcia, M., Van Damme, E., and Potters, J. (2013). Lying about what you know or about what you do? *Journal of the European Economic Association*, *11*(5), 1204-1229

Steiner, I. (2012, August 9). eBay Expands SquareTrade Warranties in Home and Garden. EcommerceBytes. https://www.ecommercebytes.com/cab/abn/y12/m08/i09/s04

Tetlock, P. E. (1985). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly*, 227-236.

Vanberg, C. (2008). Why do people keep their promises? An experimental test of two explanations 1. *Econometrica*, *76*(6), 1467-1480.

Appendix

Appendix A: Screenshots of the Z-tree program

Figure A1: Buyer's product purchase decision screen (control)

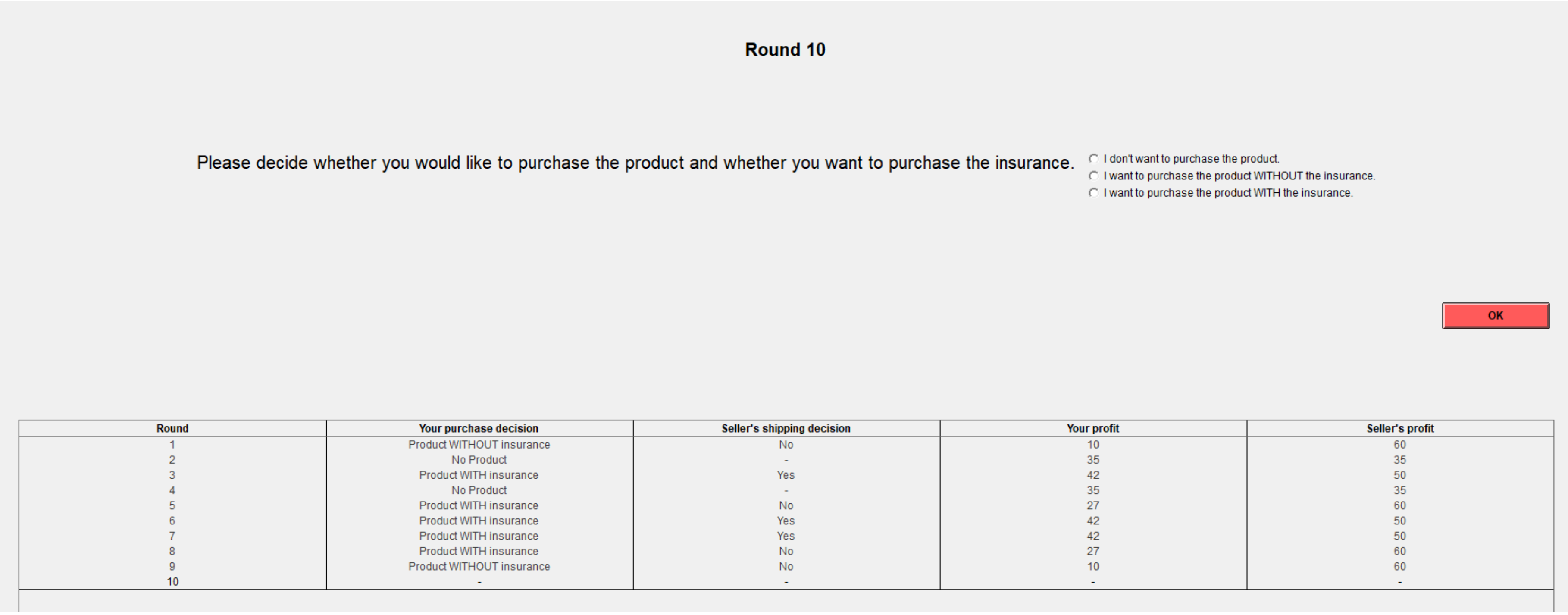


Figure A2: Seller's shipping decision screen (control)

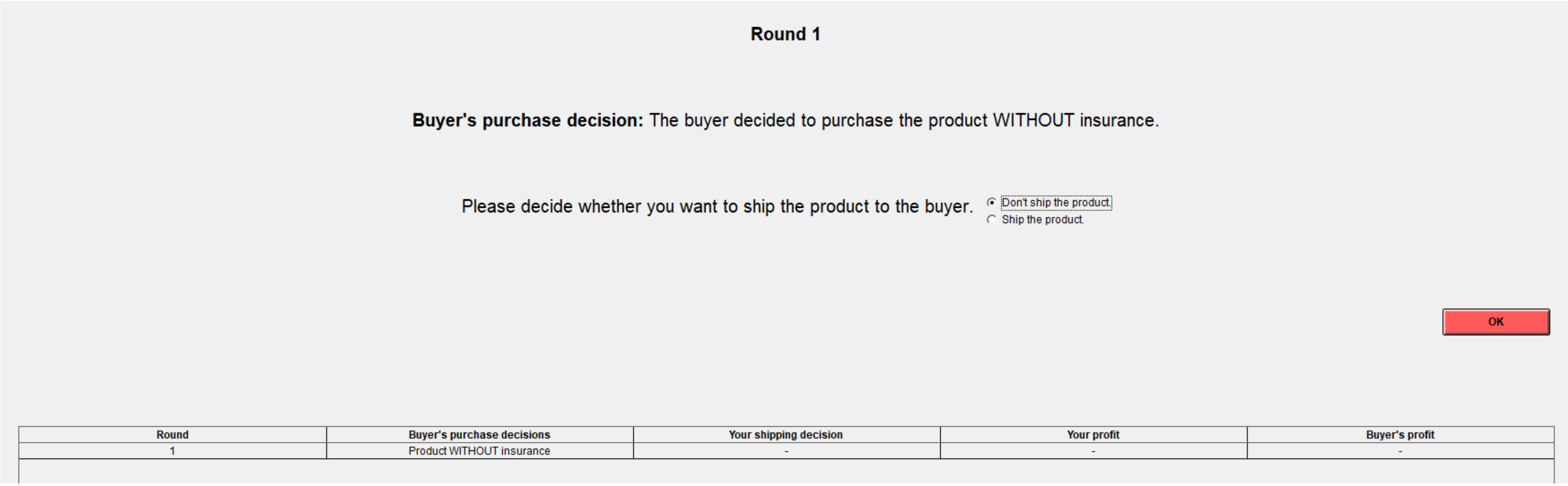
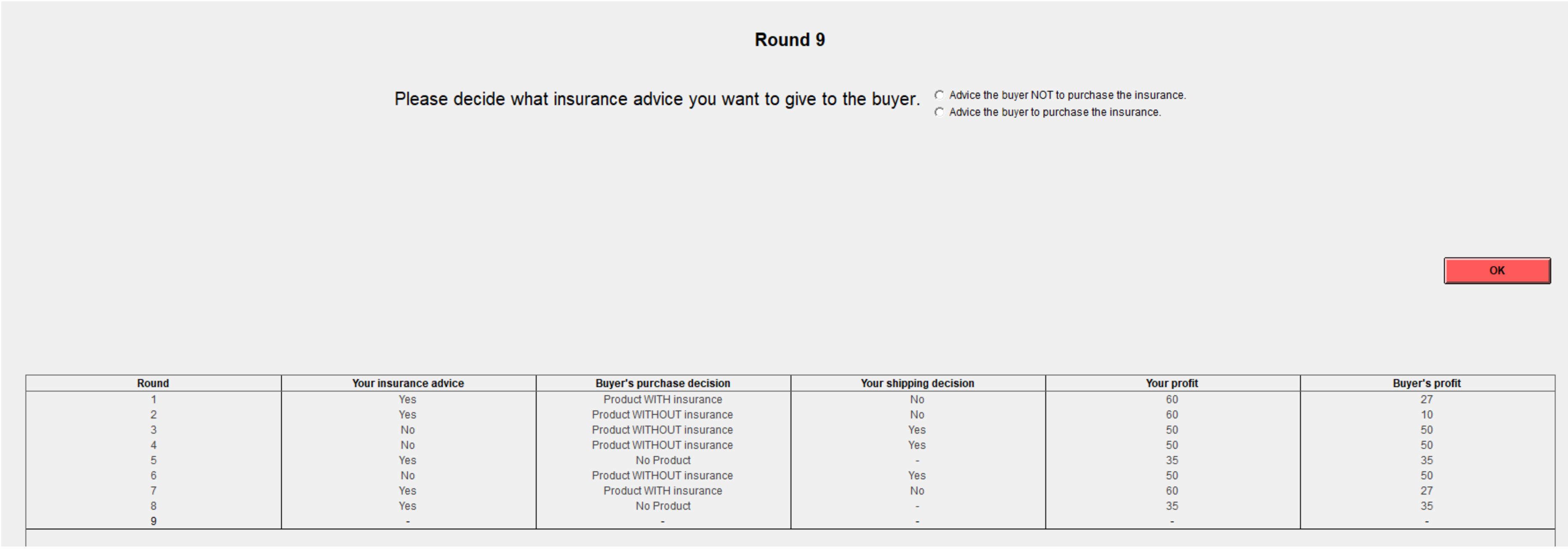


Figure A3: Seller's advice decision screen (IA treatment)



Round 9

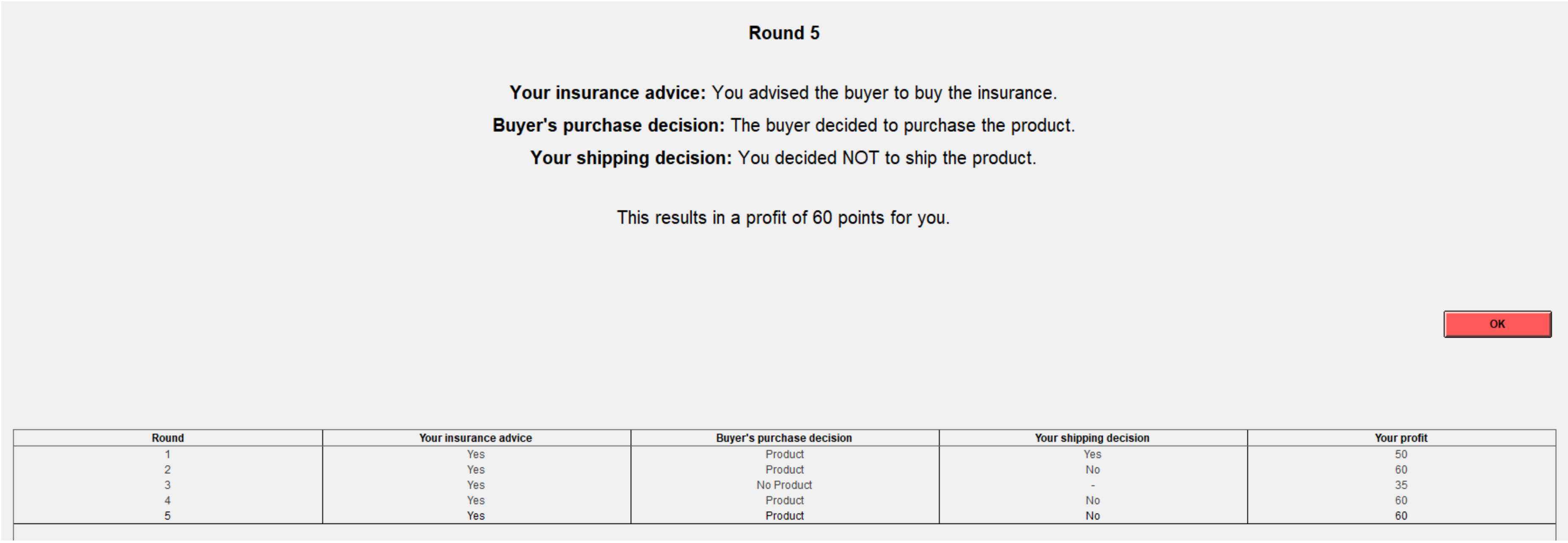
Please decide what insurance advice you want to give to the buyer.

- Advice the buyer NOT to purchase the insurance.
- Advice the buyer to purchase the insurance.

OK

Round	Your insurance advice	Buyer's purchase decision	Your shipping decision	Your profit	Buyer's profit
1	Yes	Product WITH insurance	No	60	27
2	Yes	Product WITHOUT insurance	No	60	10
3	No	Product WITHOUT insurance	Yes	50	50
4	No	Product WITHOUT insurance	Yes	50	50
5	Yes	No Product	-	35	35
6	No	Product WITHOUT insurance	Yes	50	50
7	Yes	Product WITH insurance	No	60	27
8	Yes	No Product	-	35	35
9	-	-	-	-	-

Figure A4: Seller's outcome summary screen after decisions (IA_HI treatment)



Round 5

Your insurance advice: You advised the buyer to buy the insurance.

Buyer's purchase decision: The buyer decided to purchase the product.

Your shipping decision: You decided NOT to ship the product.

This results in a profit of 60 points for you.

OK

Round	Your insurance advice	Buyer's purchase decision	Your shipping decision	Your profit
1	Yes	Product	Yes	50
2	Yes	Product	No	60
3	Yes	No Product	-	35
4	Yes	Product	No	60
5	Yes	Product	No	60

Appendix B: Instructions

(All treatments)

GENERAL INSTRUCTIONS FOR PARTICIPANTS

You are taking part in an economic experiment in which you can earn money. Please read the following instructions carefully. Your earnings depend on your decisions and on the decisions of another participant. At the end of the experiment, the amount of money earned will be paid to you in cash. Additionally, you will receive a show-up fee of 4 AUD.

Throughout the experiment, monetary amounts are not quoted in AUD, but points. Eventually, the amount of money earned during the experiment will be converted into Euro, where:

1 Point = 0.4 AUD

In this experiment, there are two types of participants, **buyers and sellers**, who make different decisions. You will only get to know your type shortly before the start of the experiment. The types will be randomly assigned and kept throughout the experiment. Please read the instructions about the decisions of both types carefully. All participants receive the same instructions.

Talking is not permitted during the experiment. Failure to comply will result in exclusion from the experiment and the loss of all earnings. If you have any questions, please address them to us: raise your hand and an experimenter will come to you.

THE EXPERIMENT

The experiment consists of 10 rounds. At the beginning of each round participants are endowed with 35 points and randomly matched in **groups of two**. Each pair consists of one buyer and one seller. At the end of the experiment only one of the 10 rounds will be randomly chosen to determine the earnings from the experiment. All rounds are equally likely to be chosen.

Product Purchase

The buyer decides whether to purchase a product from the seller at a price of 25 points.

- If the buyer chooses <u>not to purchase</u> the product, the decision task is over. The buyer and the seller keep their 35 points endowment and do not make any further decisions.
- If the buyer chooses <u>to purchase</u> the product, the seller decides whether to ship the product. Receiving the product is worth 40 points to the buyer while not receiving the product does not yield any points to the buyer. Before the seller makes the shipping decision, the buyer can decide whether to purchase an insurance.

Insurance Purchase

The insurance makes sure that the buyer is refunded the price of 25 points in case the seller does not ship the product. If the seller ships the product, the buyer will not be compensated by the insurance as no monetary loss is incurred.

The costs for the insurance are 8 points. Irrespective of whether the seller ships the product, if the insurance is purchased, the buyer needs to pay the insurance fee of 8 points.

The buyer's earnings in each round are calculated as follows:

<u>Without insurance</u>:

If the product is shipped: 35 (endowment) + 40 (the product value) - 25 (the price paid to the Seller) = 50 points.

If the product is not shipped: 35 (endowment) + 0 (no product value) - 25 (the price paid to the Seller) = 10 points.

<u>With insurance</u>:

If the product is shipped: 35 (endowment) + 40 (the product value) - 25 (the price paid to the Seller) – 8 (the insurance fee) = 42 points.

If the product is not shipped: 35 (endowment) + 0 (no product value) - 25 (the price paid to the Seller)- 8 (the insurance fee) + 25 (the refund by the insurance) = 27 points.

(IA treatment and IA_HI treatments)

Insurance Advice

Before the buyer decides whether to purchase the product, the seller is asked to give insurance advice to the buyer. The seller can advise the buyer to either not purchase or purchase the insurance. The advice is transmitted to the buyer who subsequently makes the insurance and product purchase decision.

(IA_HI treatment)

Sellers, however, will NOT know whether the buyer decided to purchase the insurance or not during or after the experiment.

Product shipping

The seller's cost for shipping the product is 10 points. If the seller does not ship the product he/she does not incur any costs. The seller's earnings in each round are calculated as follows:

- If the seller chooses to ship the product: 35 (endowment) + 25 (price paid by the buyer) – 10 points (the cost of shipping the product) = 50 points
- If the seller chooses to not ship the product: 35 (endowment) + 25 (price paid by the buyer) – 0 points (no shipping costs) = 60 points

In each round, the decisions will be made in the following order:

You are randomly divided into buyer-seller pairs. Each round proceeds in the following order:

(Control treatment)

1. Buyer: You decide whether to purchase the product. If you choose to not purchase the product the respective round is over. If you choose to purchase the product you further decide whether you want to purchase the insurance.

2. Seller: You are informed about whether the buyer has chosen to purchase the product and if yes, whether the buyer has purchased the insurance. You then decide whether to ship the

product.

3. Buyer: You are informed about the seller's shipping decision.

(IA treatment and IA_HI treatment)

1. Seller: You advise the buyer whether to purchase the insurance.

2. Buyer: You are informed about the seller's advice.

3. Buyer: You decide whether to purchase the product. If you choose to not purchase the product the respective round is over. If you choose to purchase the product you further decide whether you want to purchase the insurance.

(IA treatment)

1. Seller: You are informed about whether the buyer has chosen to purchase the product and if yes, whether the buyer has purchased the insurance. You then decide whether to ship the product.

2. Buyer: You are informed about the seller's shipping decision.

(IA_HI treatment)

1. Seller: You are informed about whether the buyer has chosen to purchase the product. You will NOT know whether the buyer has purchased the insurance. You then decide whether to ship the product.

2. Buyer: You are informed about the seller's shipping decision.

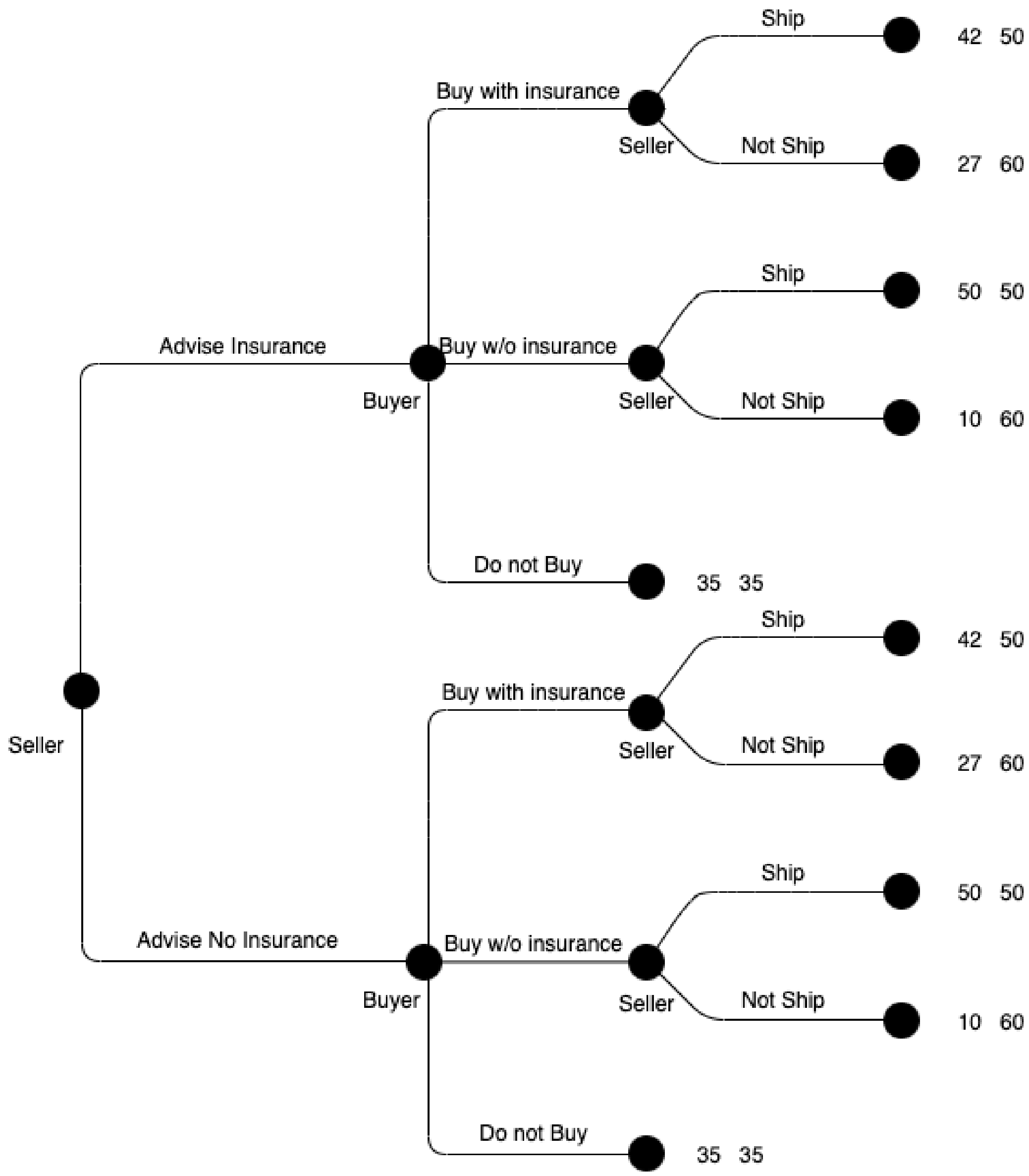
(All treatments)

After each round, buyers and sellers are randomly matched with another seller and buyer respectively. The procedure is repeated 10 times.

At the end of the experiment, participants will receive all earnings information from all rounds of the experiment and get to know which round was randomly chosen to be relevant for payment. We kindly ask you to then remain seated until you are called.

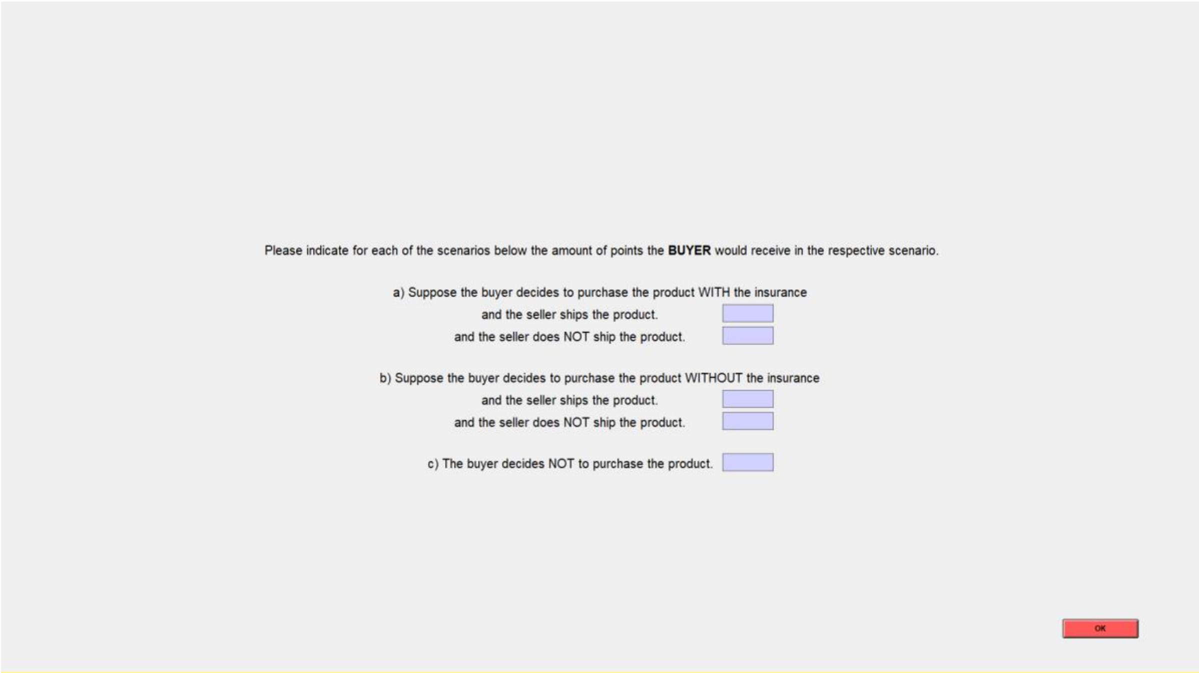
Appendix C: Buyer-Seller game with insurance advice

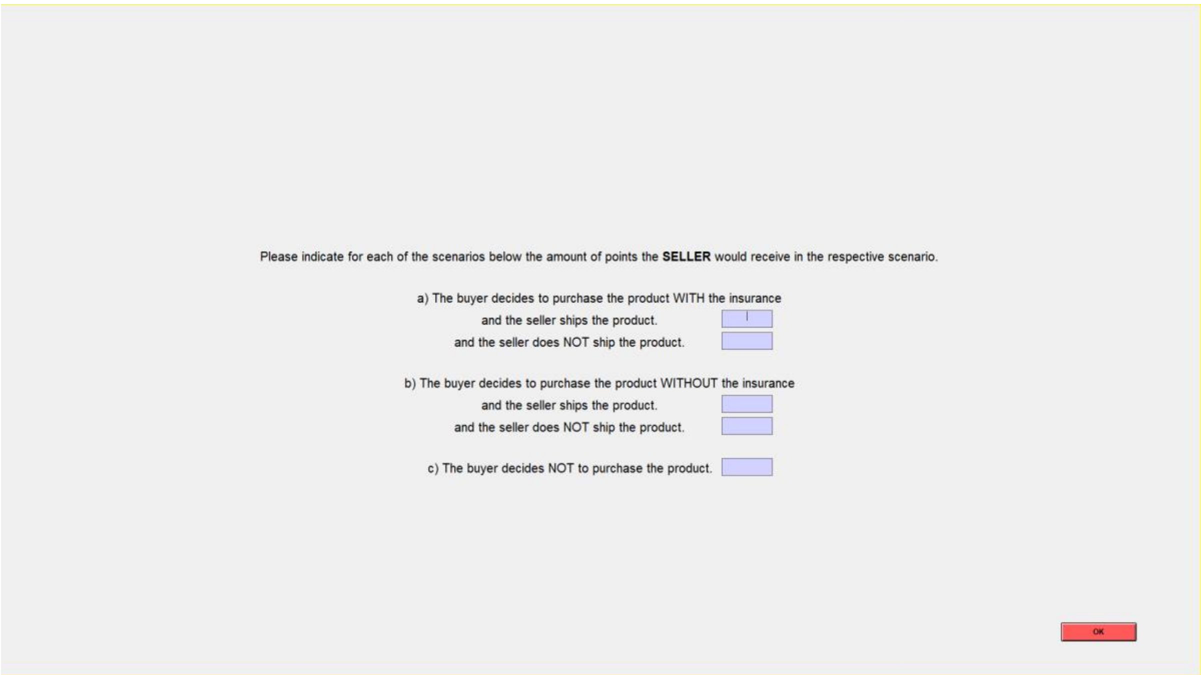
Figure C1: Buyer-seller game with insurance advice



Appendix D: Comprehension quiz screenshots.

D1: Comprehension questions in the control treatment





D2: Comprehension questions in the IA and the IA_HI treatments.

Please indicate for each of the scenarios below the amount of points the **BUYER** would receive in the respective scenario.

Suppose the seller advises the buyer NOT to buy the insurance.

a) The buyer decides to purchase the product WITH the insurance
and the seller ships the product.
and the seller does NOT ship the product.

b) The buyer decides to purchase the product WITHOUT the insurance
and the seller ships the product.
and the seller does NOT ship the product.

c) The buyer decides NOT to purchase the product.

Suppose the seller advises the buyer to buy the insurance.

a) The buyer decides to purchase the product WITH the insurance
and the seller ships the product.
and the seller does NOT ship the product.

b) The buyer decides to purchase the product WITHOUT the insurance
and the seller ships the product.
and the seller does NOT ship the product.

c) The buyer decides NOT to purchase the product.

OK

Please indicate for each of the scenarios below the amount of points the **SELLER** would receive in the respective scenario.

Suppose the seller advises the buyer NOT to buy the insurance.

a) The buyer decides to purchase the product WITH the insurance
and the seller ships the product.
and the seller does NOT ship the product.

b) The buyer decides to purchase the product WITHOUT the insurance
and the seller ships the product.
and the seller does NOT ship the product.

c) The buyer decides NOT to purchase the product.

Suppose the seller advises the buyer to buy the insurance.

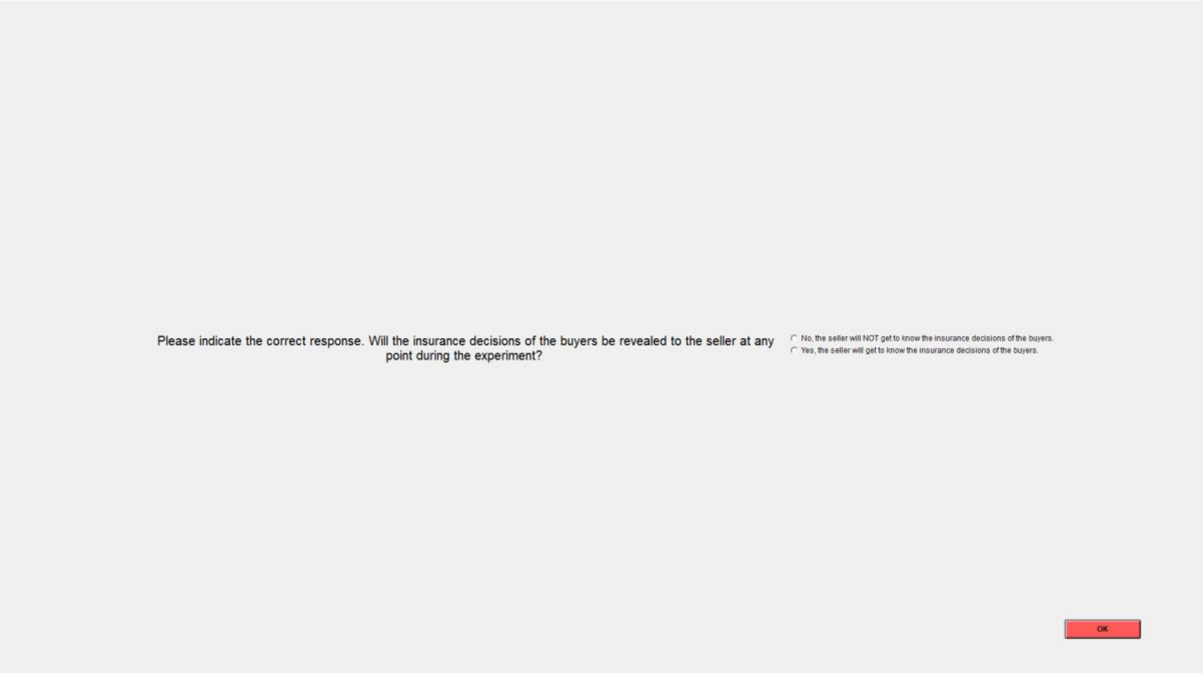
a) The buyer decides to purchase the product WITH the insurance
and the seller ships the product.
and the seller does NOT ship the product.

b) The buyer decides to purchase the product WITHOUT the insurance
and the seller ships the product.
and the seller does NOT ship the product.

c) The buyer decides NOT to purchase the product.

OK

D3: Additional comprehension question in the IA_HI treatment.



Appendix E: Omitted details of the model

E1: Theoretical analysis of the control treatment

Consider the control treatment in which the seller cannot give insurance advice to the buyer before the purchase decision is made. The proposition below characterizes the equilibrium.

Proposition 1.

In the control treatment, buyers' purchase decisions are summarized by (1). The type-g seller always ships the product, while the type-s seller never ships the product.

Proof of Proposition 1.

The type-g seller always ships the product, while the type-s seller never ships the product, given that shipping does not increase revenue but is costly. The buyer's expected payoff is $q_g v - p$ if purchasing the product without insurance, $q_g(v-p) - w$ if purchasing the product with insurance, and 0 if not purchasing the product. The buyer purchases the product without insurance only if $q_g \geq \max\left\{1 - \frac{w}{p}, \frac{p}{v}\right\}$; purchases the product with insurance if $\frac{w}{v-p} \leq q_g < 1 - \frac{w}{p}$; and does not purchase the product if $q_g < \min\{\frac{p}{v}, \frac{w}{v-p}\}$. Under the assumption $w \leq p\left(1 - \frac{p}{v}\right)$, we obtain the expression in (1). ■

Note that, if the assumption $w \leq p\left(1 - \frac{p}{v}\right)$ is violated, the buyer's payoff from purchasing the product with insurance is negative when $q_g < \frac{w}{v-p}$, and is lower than that of purchasing the product without insurance when $q_g \geq \frac{w}{v-p}$. That is, the buyer never buys insurance if $w > p\left(1 - \frac{p}{v}\right)$.

Our framework also provides insights into how the behaviour of the sellers and buyers changes with the insurance premium w, product price p, or delivery cost d. Since the type-s seller does not ship the product, a change in w, p, or d has no impact on shipping rates. However, an increase in w decreases the possibility for buyers to purchase the product without insurance, or not to purchase the product at all and thus increases the possibility for them to purchase the product with insurance. An increase in the product price p increases consumers' possibility of not purchasing the product, decreasing their possibility of purchasing the product without insurance and has an ambiguous effect on the possibility of them purchasing the product with insurance. The delivery cost d does not enter the buyer's decision-making.

E2: Theoretical analysis of the IA treatment

Let $\sigma_g(i) \in [0,1]$ denote the probability with which the buyer believes that the seller is of type-g after receiving advice $i = N, Y$.

Proposition 2.

Suppose the seller can advise the buyer whether to buy the insurance before he makes the product purchase decision.

- There always exists an *N*-pooling equilibrium in which all sellers advise *N*, provided the off-path belief is such that $\sigma_g(Y) < \frac{w}{v-p}$
  - When $d \leq \alpha$, all sellers ship the product upon receiving orders. Buyers purchase the product without insurance when receiving advice *N*, and do not purchase the product when receiving advice *Y*.
  - When $d > \alpha$, upon receiving orders, the type-g seller ships the product, and the type-s seller does not ship the product. Buyers behave the same way as in the control treatment when receiving advice *N*, and do not purchase the product when receiving advice *Y*.
- There also exists a *Y*-pooling equilibrium in which all sellers advise *Y* if $q_g \geq \frac{w}{v-p}$, or if $q_g < \frac{w}{v-p}$ and $d > \alpha$, provided the off-path belief is such that $\sigma_g(N) < \frac{w}{v-p}$.
  - When $q_g \geq \frac{w}{v-p}$, upon receiving orders, the type-g seller ships the product, and the type-s seller does not ship the product. Buyers behave the same way as in the control treatment when receiving advice *Y*. The buyer's choice can be arbitrary when receiving *N*.
  - When $q_g < \frac{w}{v-p}$ and $d > \alpha$, upon receiving orders, the type-g seller ships the product, and the type-s seller does not ship the product. The buyer does not purchase the product after receiving any advice.

Proof of Proposition 2.

Consider first the N-pooling equilibrium. Both types of sellers advise N on the equilibrium path. If the buyer follows the advice and purchases the product without insurance, the type-s seller's profit is $p - d$ if she delivers the product, and $p - \alpha$ if she does not deliver the product. She will deliver the product, provided $d \leq \alpha$. Knowing $d \leq \alpha$, the buyer will purchase the product without insurance when receiving advice N. If receiving advice Y, the buyer believes that the seller is of type-g with a probability $\sigma_g(Y)$. He also understands that a type-s seller will not ship the product after advising Y. The analysis of the buyer's choice is the same as in the control after replacing q_g by $\sigma_g(Y)$. So, we know that the buyer will not purchase the product if $\sigma_g(Y) < \frac{w}{v-p}$, implying the type-s seller does not want to deviate to advising Y if $\sigma_g(Y) < \frac{w}{v-p}$.

Now suppose $d > \alpha$. The type-s seller does not ship the product even after advising N. Upon observing N, the buyer keeps his prior and believes the seller is of type-g with a probability q_g. Given that the type-s seller never delivers the product when $d > \alpha$, the buyer's choice is the same as in the control treatment. The buyer will not purchase the product once receiving advice Y as the off-path belief is such that $\sigma_g(N) < \frac{w}{v-p}$.

Next, consider the Y-pooling equilibrium. In this equilibrium, both types of sellers advise Y on the equilibrium path. Since then the psychological cost never arises, the buyer behaves in the same way as in the control. Then, consider whether the type-s seller wants to deviate to advising N. Note that when $q_g < \frac{w}{v-p}$, buyers do not buy after receiving Y, and the type-s seller's profit is zero. If she instead advises N, the buyer should believe that the product will be delivered if $d \leq \alpha$. Then, the type-s seller's profit is $p - d > 0$. Thus, the Y-pooling equilibrium does not exist if $q_g < \frac{w}{v-p}$ and $d \leq \alpha$. For the remaining range of parameters, the Y-pooling equilibrium exists as the type-s seller does not want to deviate to advising N. If $q_g \geq \frac{w}{v-p}$, the type-s seller makes a profit equal to p by advising Y and will never deviate. If $q_g < \frac{w}{v-p}$ but $d > \alpha$, the type-s seller always receives a zero profit no matter what she advises since the off-path belief is such that $\sigma_g(N) < \frac{w}{v-p}$. ■

We now explain how sellers' and buyers' behaviour is affected by the parameters of the model in the IA treatment. Provided $d > \alpha$, the impacts of changes in p or w on the purchasing and shipping decisions are the same as in the control. Now suppose $d \leq \alpha$. All sellers advise

N and ship the product, and all buyers purchase the product without insurance. So, again, the purchasing and shipping decisions are unaffected by a change in p or w. Whether sellers ship the product depends on the comparison between the shipping cost d and the psychological cost α. A change in d does not affect any behaviour as long as it does not reverse the relation between d and α. However, a substantial increase in d, which reverses $d \leq \alpha$ to $d > \alpha$, will nullify the insurance advice mechanism, and decrease both the purchasing and shipping rates.

We next show that there does not exist a separating equilibrium. Consider a separating equilibrium, in which on the equilibrium path a type-g buyer advises *N*, while a type-s seller advises *Y*. Note that a type-s seller never incurs the psychological cost when advising *Y*, irrespective of what the buyer does. As a result, a type-s seller will never ship the product given that delivery is costly. Therefore, the buyer will not purchase the product upon receiving the advice *Y*. If he purchases the product without insurance, his payoff is $-p < 0$. If he purchases the product with insurance, his payoff is $-w < 0$. Anticipating that the buyer will not purchase the product, a type-s seller will want to deviate to advising *N* instead. Therefore, there does not exist a separating equilibrium in which a type-g seller advises *N* and a type-s seller advises *Y*. Moreover, there does not exist a separating equilibrium in which a type-g seller advises *Y* and a type-s seller advises *N*. Note that a type-s seller can be strictly better by deviating to advising *Y*. By doing so, the type-s seller pretends to be a type-g seller, thereby increasing the probability of making sales, but never incurring the psychological cost.

There does not exist a semi-separating equilibrium except for a trivial case, in which there is no transaction between the buyer and the seller. Note that by advising *Y*, the type-s seller reveals her type and the buyer will not purchase the product. This implies the profit of advising *Y* is zero for the type-s seller. Since the type-s seller randomizes between *N* and *Y*, her expected profit from advising *N* must be also zero. Below we show this cannot be the case. If $d \leq \alpha$, the type-s seller will deliver after advising *N*. The buyer's payoff is $v - p > 0$ if he only purchases the product, $-w$ if he purchases the product with insurance, and zero if he does not purchase the product. So, the buyer will purchase the product only. The type-s seller's expected profit from advising *N* is $p - d > 0$. Therefore, there does not exist a semi-separating equilibrium when $d \leq \alpha$ as the type-s seller's profit from advising *N* is strictly positive and therefore will not randomize. If $d > \alpha$, the type-s seller will not deliver the product no matter what the buyer does. The buyer's behavior will be the same as in the control. The type-s seller's expected profit is either p (i.e., the buyer purchases the product with the insurance), $p - \alpha$

(i.e., the buyer only purchase the product), or 0 if the buyer does not purchase the product. So, the type-s seller randomizes only in the last case when there is no transaction at all.

The *Y*-pooling equilibrium can, however, be ruled out by applying refinement using forward induction. To utilize forward induction, assume that type-g sellers (i) prefer buyers purchasing the product to not purchasing the product when the price is fixed and the product will be shipped with probability one, and (ii) prefer buyers not to buy the insurance if shipping choices are fixed. Suppose in equilibrium both types of sellers advise *Y*. Imagine that buyers observe the off-path advice *N*. We claim that buyers should believe the seller is of type g. This is because, any buyer response that can induce a type-s seller to deviate to advise *N* will also induce a type-g seller to deviate, while the reverse is not true, since type-s sellers can potentially incur a psychological cost. Thus, the set of buyer responses that induce type-g sellers to advise *N* is strictly larger than the set of buyer responses that induce type-s sellers to advise *N*. Knowing buyers' responses upon receiving *N*, both types of sellers will deviate to advise *N* when $q_g < \frac{w}{v-p}$. Type-g sellers will deviate to advise *N* when $q_g \geq \frac{w}{v-p}$. These profitable deviations break the *Y*-pooling equilibrium.

In contrast, in an *N*-pooling equilibrium, the set of buyer responses that can induce a type-g seller to advise *Y* is null as a type-g seller reaches its maximum profit on the equilibrium path. Thus, upon receiving *Y*, buyers must believe that the seller is a type-s seller and therefore do not purchase the product. This implies the *N*-pooling equilibrium survives forward induction.

In our experiment, we set up w=8, p=25, and v=40 (thus, $\frac{w}{v-p}$ =53%) such that $q_g > \frac{w}{v-p}$ can only happen when more than half of the sellers are non-strategic (>53%). In our control treatment, we find the proportion of type-g sellers who always ship the product is only 18.5%, much smaller than 53%. Thus, we may argue that the *N*-pooling equilibrium is likely to be the unique equilibrium in the experiment.

Comparison between control and IA treatments

When $d > \alpha$, the equilibrium outcome remains the same as in the control treatment. Thus, we focus on the condition $d \leq \alpha$. In the IA treatment, sellers advise *N*. In the control treatment, only buyers with $q_g \geq \frac{w}{v-p}$ purchase the product and among them those with $q_g < 1 - \frac{w}{p}$ also buy insurance. In the IA treatment, all buyers purchase the product without purchasing insurance. In the control treatment, the type-s seller never ships the product. In the IA treatment, all sellers ship the product.

E3: Theoretical analysis of IA_HI treatment

In the IA_HI treatment, we select the *N*-pooling equilibriam whenever there co-exist both *N*-pooling and *Y*-pooling equilibrim for the same reason as in Proposition 2, and select the mixed-strategy equilibrium whenver there co-exist both pure-strategy and mixed-strategy equilibria as it requires less coordination between buyers and sellers when buyers' insuarance purchasing decision is unobservable by the seller. After applying this equilibrium selection rule, the following proposition characterizes the equilibrium in the IA_HI.

Proposition 3.

Suppose that the seller can provide insurance advice but she cannot observe whether the buyer buys the insurance. Both types of sellers advise *N*. Buyers' off-path belief when receiving an advice Y is that the seller is a type-g seller with probablity $\sigma_g(Y) < \frac{w}{v-p}$. The remainder of equilibrium characterization is given as follows:

- Suppose $d > \alpha$. Upon receiving orders, the type-g seller ships the product, and the type-s sellers does not ship. Buyers behave the same way as in the control treatment when receiving advice *N*, and do not purchase the product when receiving advice *Y*.
- Suppose $d \leq \alpha$ and $q_g \geq 1 - \frac{w}{p}$. Upon receiving orders, both types of sellers ship the product. Buyers buy the product when receiving advice *N*, and do not purchase the product when receiving advice *Y*.
- Suppose $d \leq \alpha$and $q_g < 1 - \frac{w}{p}$. Upon receiving *N*, buyers purchase the product without insurance with a probability $\frac{d}{\alpha}$ and purchase the product with insurance with a probability of $1 - \frac{d}{\alpha}$. Upon receiving orders, the type-g seller always ships the product, while the type-s seller ships the product with a probability of $1 - \frac{w}{p(1-q_g)}$ and does not ship the product with a probability of $\frac{w}{p(1-q_g)}$. Upon receiving advice *Y*, buyers do not purchase the product.

Proof of Proposition 3.

In the IA_HI treatment, it is clear that the insurance advice mechanism cannot work if the psychological cost is not sufficiently high, i.e., $d > \alpha$, Then, everything remains the same as in the control. Let us focus on the case when $d \leq \alpha$.

Firstly, consider the case when $q_g \geq 1 - \frac{w}{p}$. Note that for this range of parameters, buyers purchase the product without insurance in the control treatment. A type-s seller does not want to advise *Y* as that will reveal her type. Buyers continue to hold the same prior belief after observing *N* and purchase the product without insurance. The type-s seller is better off shipping the product as she will incur the psychological cost otherwise.

Secondly, consider the case when $q_g < 1 - \frac{w}{p}$. We denote γ as buyers' probability of buying without insurance and δ as the type-s seller's probability of shipping the product. The buyer's expected payoff is $[q_g + (1 - q_g)\delta]v - p$ if he follows the advice of *N* and purchases the product without insurance, and $[q_g + (1 - q_g)\delta](v - p) - w$ if he does not follow the advice and purchases the product with insurance. The buyer chooses $\gamma \in (0,1)$ if he is indifferent between the choices, which implies $\delta = 1 - \frac{w}{p(1-q_g)}$. Our assumption $w \leq p\left(1 - \frac{p}{v}\right)$ ensures this mixed-strategy equilibrium exists as otherwise buying both product and insurance is a strictly dominated option. In this mixed-stategy equilibrium, a type-s seller's expected profit is $p - d$ if she ships the product and $p - \gamma\,\alpha$ if she does not ship. She randomizes between the choices only if she is indifferent, which implies $\gamma = \frac{d}{\alpha}$.

If the off-path advice *Y* is observed, the buyer believes the seller is of type-g with a probability $\sigma_g(Y)$. He also understands that a type-s seller will never ship the product after advising *Y* as the psychological cost does not arise. His payoff is $\sigma_g(Y)v - p$ if he purchases the product only, and is $\sigma_g(Y)(v - p) - w$ if he also buys the insurance. The buyer does not purchase the product if $max\{\sigma_g(Y)v - p, \sigma_g(Y)(v - p) - w\} < 0$. Given the assumption $w \leq p\left(1 - \frac{p}{v}\right)$, this is equivalent to $\sigma_g(Y) < \frac{w}{v-p}$. Thus, the buyer does not purchase the product after receiving Y if $\sigma_g(Y) < \frac{w}{v-p}$, and the type-s seller will not deviate to advising *Y*.■

Note that there does not exist a mixed-strategy equilibrium in which buyers mix between "purchase the product without insurance" and "not purchase the product". This is because for this equilibrium to exist the type-s seller must randomize between "ship" and "not ship". However, the shipping decision is conditional on purchasing the product. If the product is not purchased and given that the type-s seller can see that, the type-s seller will choose "ship" with zero probability. Consequently, this means the buyer cannot choose to "not purchase the product" with non-zero probability. For the same reason, buyers cannot randomize between "purchase the product with insurance" and "not purchase the product" in equilibrium. Also, this mixed-strategy equilibrium play only exists when the advice is N. If both N and Y can appear on the equilibrium path (i.e., the type-s seller randomizes between *N* and *Y*), and given that the type-s seller has no psychological cost following advising Y, they will choose "ship" with probability zero. This breaks the mixed-strategy equilibrium.

In the IA_HI treatment, both sides use a mixed strategy in equilibrium. The shipping rate decreases in w and increases in p, while buyers buy insurance less frequently when the shipping cost increases. It is well-known that in an equilibrium with both sides using a mixed strategy the comparative statics are often counter-intuitive as players need to adjust their strategies to keep the opponent indifferent when there is an exogenous shock.

Comparison between control, IA, and IA_HI treatments

In both the IA and IA_HI treatments, all buyers purchase the product. In the control, only buyers with $q_g \geq \frac{w}{v-p}$ purchase the product. In both the IA and IA_HI treatments, sellers advise *N*. In the IA treatment, all buyers purchase the product without insurance, while in IA_HI some buyers purchase the product with insurance as they are mixing between "purchase without insurance" and "purchase with insurance". In the IA treatment, all sellers ship the product, while in the IA_HI treatment, only some type-s sellers ship the product as they are mixing between "ship" and "not ship". In the control, however, no type-s sellers ship the product. The predictions in Hypotheses 1 to 4 thus follow.

E4

Discussions of some assumptions on the theoretical model

In our framework, we made a few assumptions. Below we discuss what happens when the assumptions are violated.

Assumption 1: Some sellers always ship and some sellers are strategic.

This assumption is consistent with the recent findings from behavioral economics literature that many people are prosocial. If instead, we assume all sellers are of type-s, this will not qualitatively change the model's predictions. To see this, suppose q_g=0 and therefore all sellers are strategic type-s sellers. Buyers never purchase the product in the control treatment as no sellers will ship. In the IA treatment, there exists an equilibrium where all sellers advise N, therefore committing to shipping the product, and all buyers purchase the product when $d \leq \alpha$. However, if $d > \alpha$, no seller will ship the product and no buyer will purchase the product. In the IA_HI treatment, the mixed-strategy equilibrium involves sellers shipping the product with probability $\delta = 1 - \frac{w}{p}$ and buyers purchase the product without insurance with probability $\gamma = \frac{d}{\alpha}$.

Assumption 2: the insurance is not too expensive, i.e., $w \leq p\left(1 - \frac{p}{v}\right)$.

This assumption means that in each treatment at least some buyers will buy insurance. Our experiment is consistent with this assumption, as noted in footnote 7. We now consider in theory what happens if the insurance is more expensive, i.e., $w > p\left(1 - \frac{p}{v}\right)$.

In the control, this means no buyer will buy the insurance as we explained in the proof of Proposition 1.

In the IA treatment, although rational buyers would not buy the insurance, the *N*-pooling equilibrium can continue to hold if $d \leq \alpha$. This is indeed the case when $\sigma_g(Y) < \frac{p}{v}$. If the advice is instead *Y*, the buyer believes this seller is of type-g with a probability $\sigma_g(Y)$. He will not purchase the product if $\sigma_g(Y)v - p < 0$, or equivalently, $\sigma_g(Y) < \frac{p}{v}$. In this case, the equilibrium outcome is the same as in the case of $w \leq p\left(1 - \frac{p}{v}\right)$ and the insurance advice mechanism still helps promote cooperative behaviour.

In the IA_HI treatment, buyers can no longer randomize between "purchasing the product with insurance" and "purchasing the product without insurance" as the former option implies a negative payoff due to $w > p\left(1 - \frac{p}{v}\right)$. The remaining equilibria are the two pure-strategy ones resulting from extreme beliefs. If sellers believe that all buyers will follow the advice *N* and purchase the product without insurance, then the equilibrium is the same as in the IA treatment, i.e., all buyers purchase the product without buying insurance, and all sellers ship the product. If sellers believe that no buyers will follow the advice *N* and will buy insurance in case they purchase the product, then in equilibrium sellers will advise *N* and both sides will behave as in the control, and in particular, no buyers purchase the product when $p_g < \frac{w}{v-p}$.

Appendix F: Regressions with control variables

F1: Random individual effects linear probability regression analysis of insurance advice decisions with controls

Independent variables	Dependent variable: Advice $N_{i,t}$ =1, if seller *i* advised *N* in round *t* =0, o.w.	
	(1)	(2)
IA	0.128***	0.0714
	(0.043)	(0.083)
Round		0.0133*
		(0.007)
IA_Round		0.0102
		(0.010)
Constant	0.650***	0.577***
	(0.068)	(0.092)
Controls	Y	Y
N	1,090	1,090

Note: IA_HI is the baseline. Controls include gender, major, and how well the individual understood the experiment instructions. Robust standard errors clustered at the session level are reported in the parentheses*** p<0.01, ** p<0.05, * p<0.1

F2: Random individual effects LPM regression analysis of product purchase decisions with controls

Independent variables	Dependent variable: Buy$_{i,t}$ =1, if buyer *i* purchased the product in round *t* = 0, o.w.					
	(1) IA and Control	(2) IA and Control	(3) IA and Control	(4) IA_HI and Control	(5) IA_HI and Control	(6) IA_HI and Control
β_1: IA	0.166***	0.054	-0.122			
	(0.057)	(0.075)	(.105)			
β_2: IA_HI				0.168**	0.034	0.010
				(0.074)	(0.073)	(0.078)
β_3: Round		-0.042***	-0.042***		-0.042***	-0.042***
		(0.007)	(0.007)		(0.007)	(0.007)
β_4: IA*Round		0.020*	-0.018			
		(0.012)	(0.153)			
β_5: IA_HI*Round					0.024**	-0.001
					(0.010)	(0.014)
β_6: Advice *N*			0.286***			0.043
			(0.079)			(0.017)
β_7: Advice *N**Round			0.033***			0.032*
			(0.009)			(0.017)
Constant	0.554***	0.787***	0.778***	0.733***	0.733***	0.732***
	(0.076)	(0.078)	(0.078)	(0.081)	(0.081)	(0.082)
H0: $\beta_3 + \beta_4$=0	p=0.026	p=0.026	p<0.001			
H0: $\beta_3 + \beta_5$=0				p=0.012	p=0.012	p<0.001
Controls	Y	Y	Y	Y	Y	Y
N	1120	1120	1120	1080	1080	1080

Note: Controls include gender, major, and how well the individual understood the experiment instructions. Robust standard errors clustered at the session level are reported in the parentheses, *** p <0.01, ** p <0.05, * p <0.1.

F3: Random individual effects LPM regression analysis of shipping decisions with controls

Independent variables	Dependent variable: $Ship_{j,t}$ =1, if the seller j shipped the product in round t;					
	(1) IA and Control	(2) IA and Control	(3) IA (advice N) and Control	(4) IA_HI and Control	(5) IA_HI and Control	(6) IA_HI (advice N) and Control
β_1: IA	0.239***	-0.002	0.139**			
	(0.050)	(0.076)	(0.057)			
β_2: IA_HI				0.111	-0.050	0.182**
				(0.070)	(0.079)	(0.081)
β_3: Noinsure			0.045			0.045
			(0.060)			(0.060)
β_4: Noinsure*IA			0.129*			
			(0.066)			
β_5: Noinsure*IA_HI						-0.015
						(0.064)
β_6: Advice *N*		0.285***			0.217***	
		(0.067)			(0.073)	
Constant	0.325***	0.344***	0.345***	0.544***	0.546***	0.516***
	(0.061)	(0.056)	(0.061)	(0.086)	(0.086)	(0.103)
H0: $\beta_3 + \beta_4 = 0$			$p < 0.001$			
H0: $\beta_3 + \beta_5 = 0$						$p = 0.216$
Controls	Y	Y	Y	Y	Y	Y
N	651	651	614	635	635	552

Note: Controls include gender, major, and how well the individual understood the experiment instructions. Robust standard errors clustered at the session level are reported in the parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix G: Other Graphs

Figure G1: Proportion of buyers who purchased the product without insurance in each round

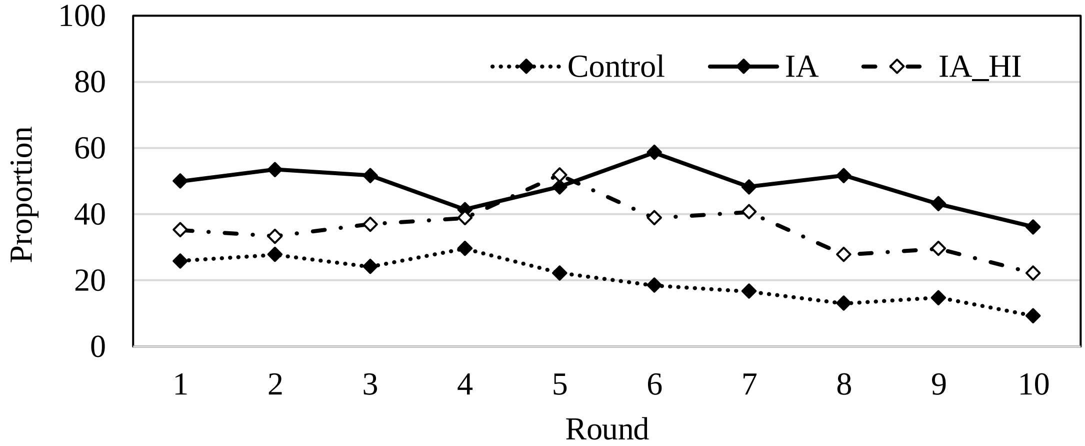
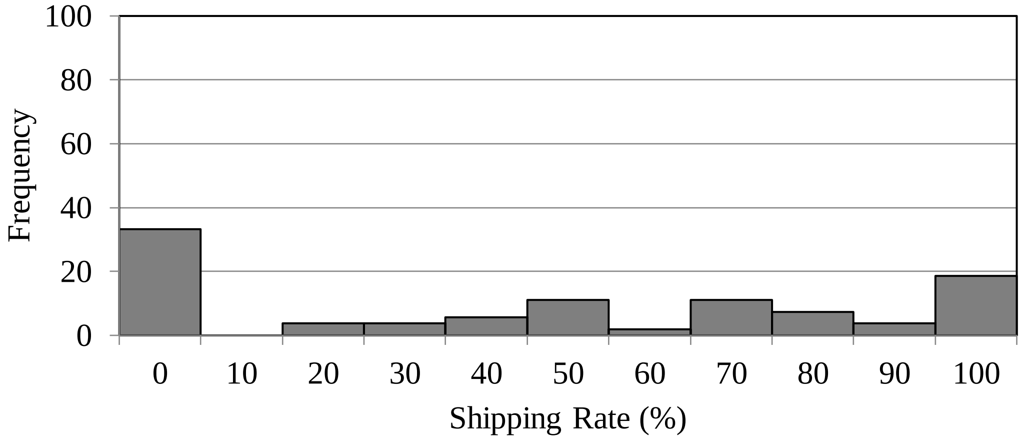


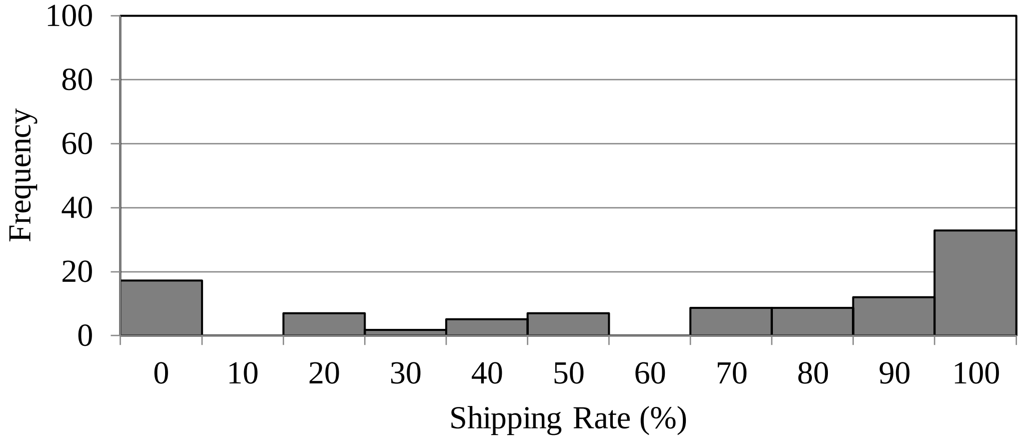
Figure G2: Distribution of shipping rate

a) Control treatment



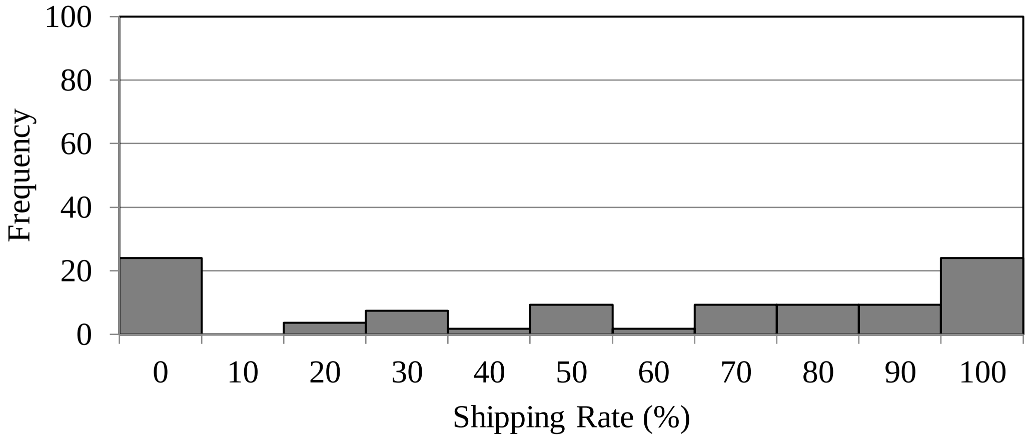
Note: # of observations=54

b) IA treatment



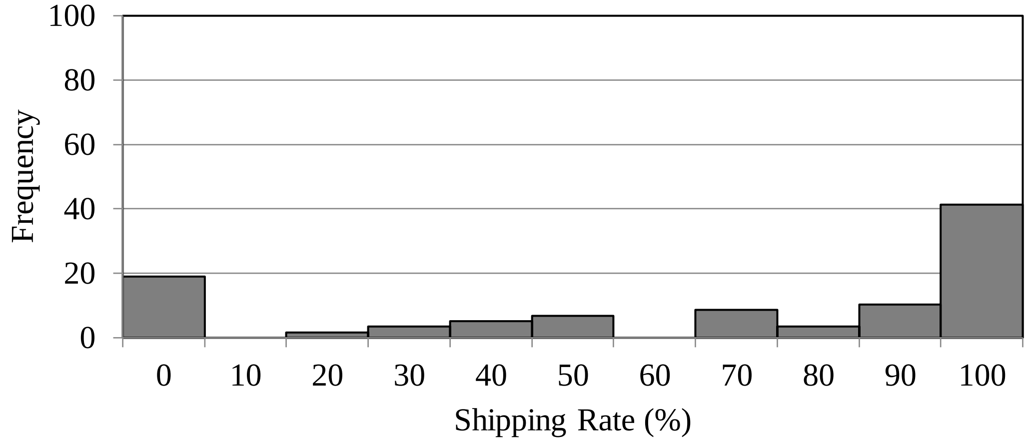
Note: # of observations=58

c) IA_HI treatment



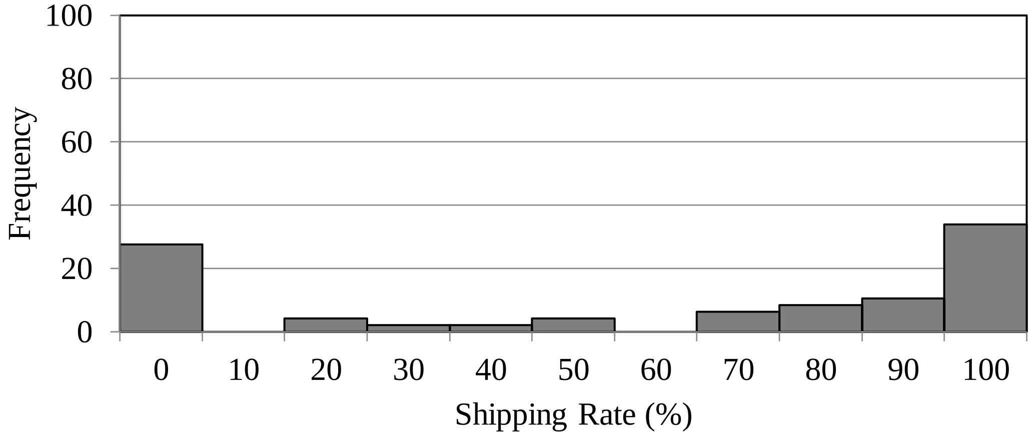
Note: # of observations=54

d) IA treatment (sellers advised *N*)



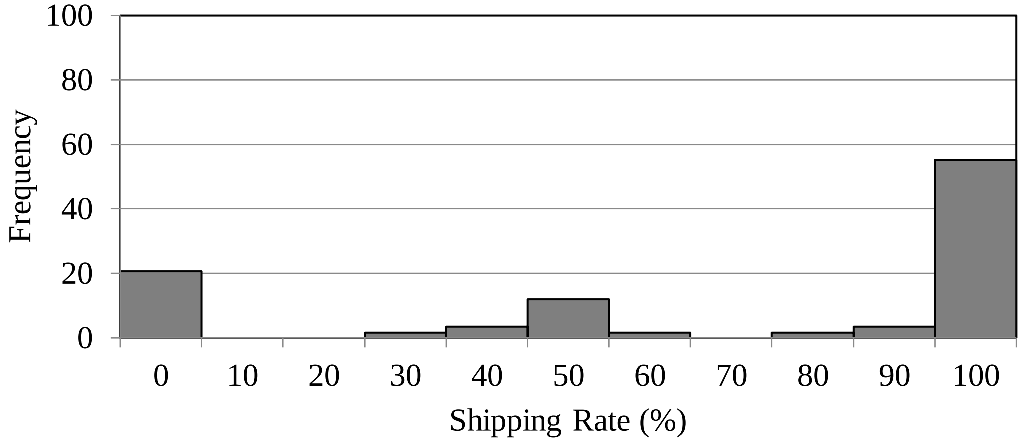
Note: # of observations=58

e) IA_HI treatment (sellers advised *N*)



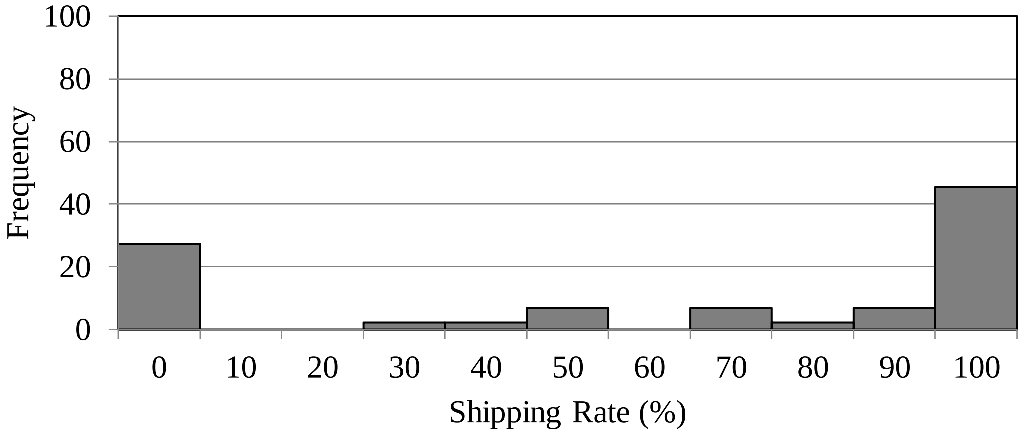
Note: # of observations=47

f) IA treatment (sellers advised *N* and buyers purchased the product without insurance)



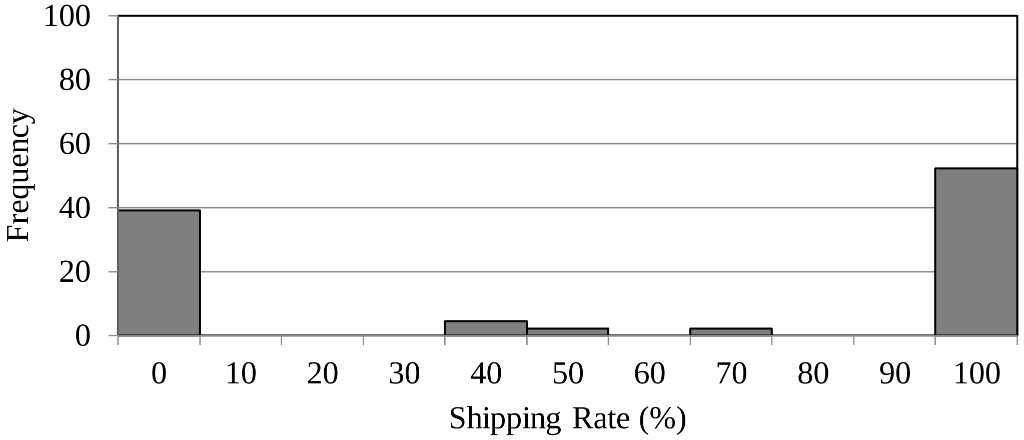
Note: # of observations=58

g) IA_HI treatment (sellers advised *N* and buyers purchased the product without insurance)



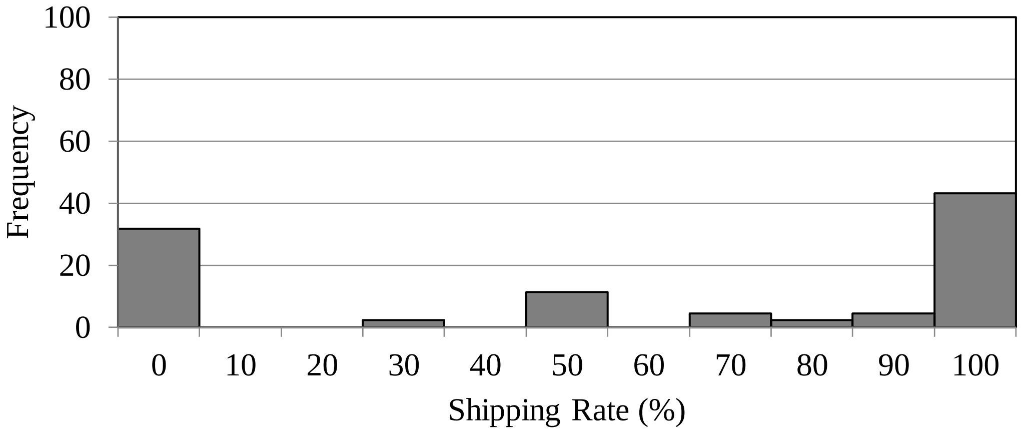
Note: # of observations=44

h) IA treatment (sellers advised *N* and buyers purchased the product with insurance)



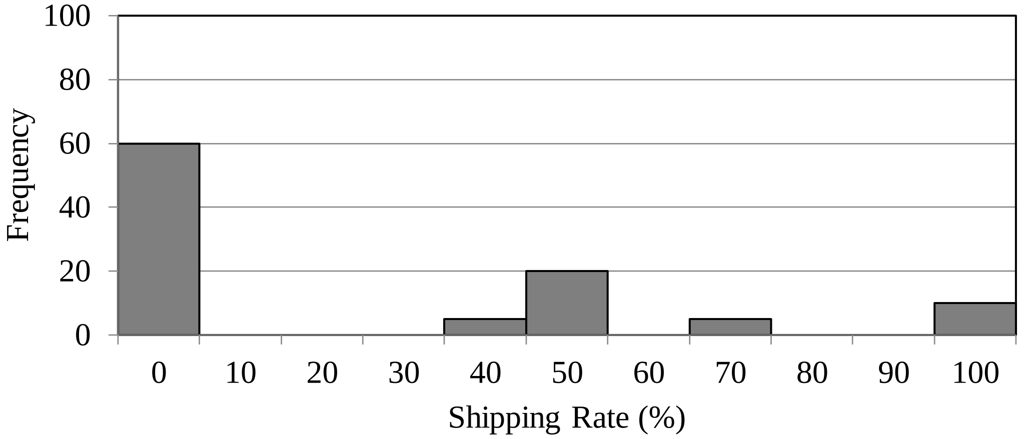
Note: # of observations=46

i) IA_HI treatment (sellers advised *N* and buyers purchased with insurance)



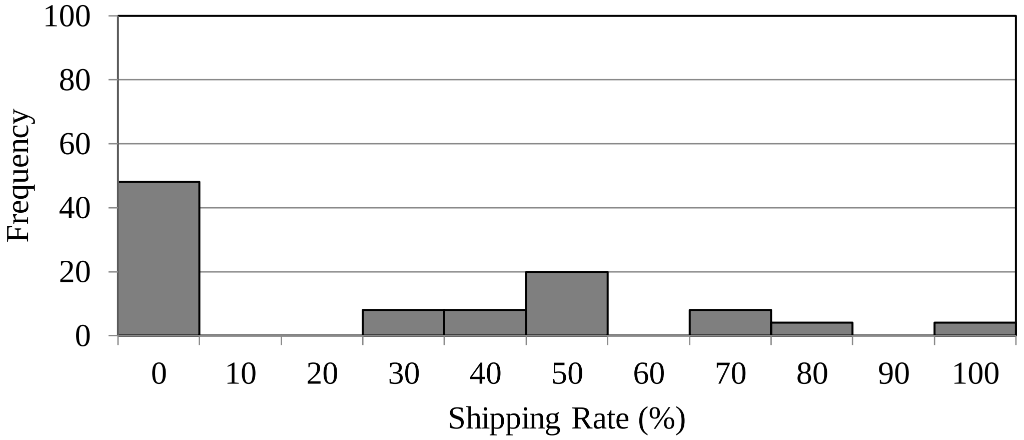
Note: # of obs.=44

j) IA treatment (sellers advised *Y*)



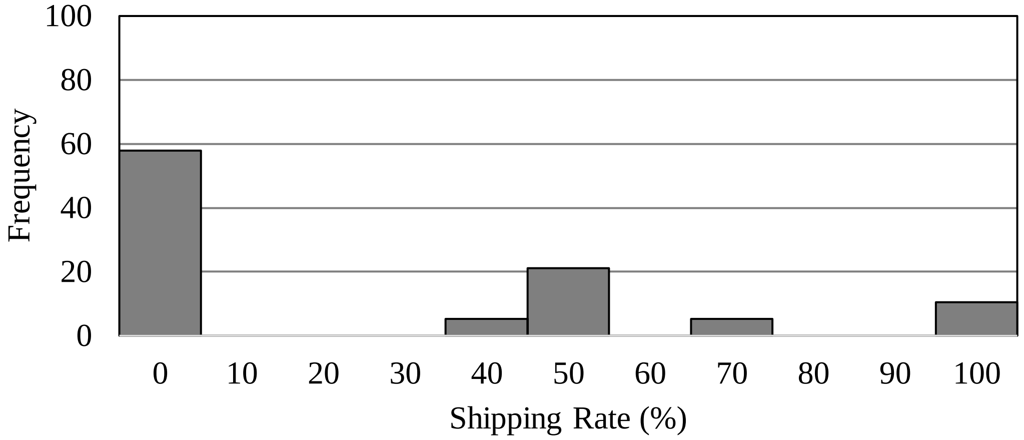
Note: # of observations=20

k) IA_HI treatment (sellers advised *Y*)



Note: # of observations=25

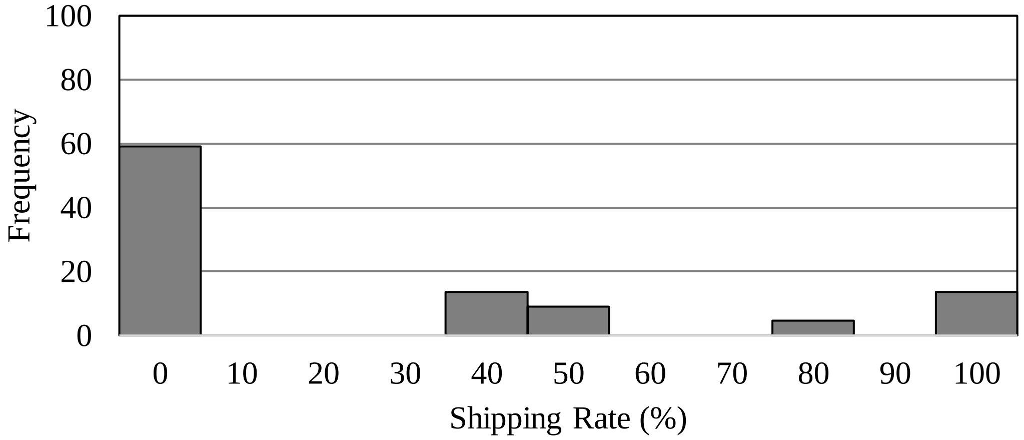
l) IA treatment (sellers advised *Y* and buyers purchased with insurance)



Note: # of observations=19.

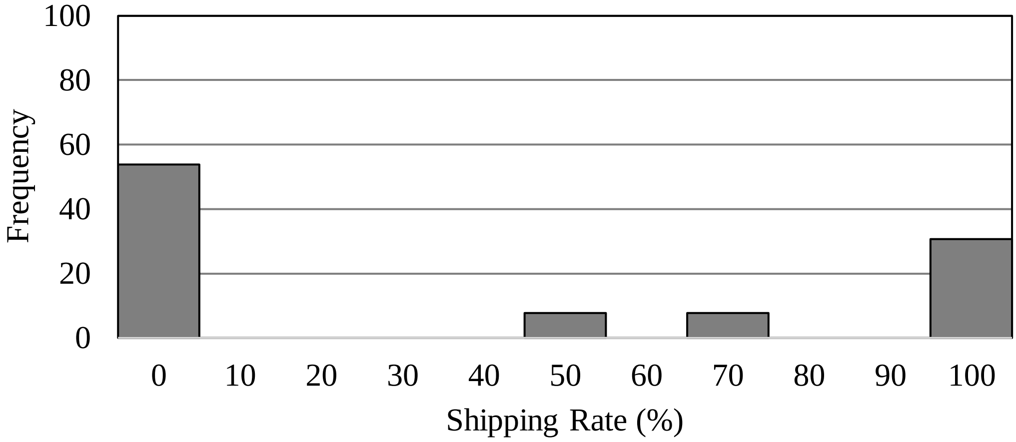
We do not include the distribution of shipping decisions for the IA treatment when sellers advised Y and buyers purchased the product without insurance because there are only four observations.

m) IA_HI treatment (sellers advised *Y* and buyers purchased the product with insurance)



Note: # of observations=22.

n) IA_HI treatment (sellers advised *Y* and buyers purchased the product without insurance)



Note: # of observations=13.

Appendix H: Probit Regressions:

H1: Random individual effects probit regression analysis of insurance advice decisions

Independent variables	Dependent variable: Advice $N_{i,t}$ =1, if seller *i* advised *N* in round *t* =0, o.w.			
	(1) Probit	(2) Marg. Eff	(3) Probit	(4) Marg. Eff.
β_1: IA	0.329**	0.100**	0.027	0.008
	(0.139)	(0.039)	(0.247)	(0.073)
β_2: Round			0.041**	0.012**
			(0.019)	(0.006)
β_3: IA_Round			0.061*	0.018*
			(0.034)	(0.010)
Constant	0.557***		0.335***	
	(0.068)		(0.125)	
N	1120	1120	1120	1120

Note: IA_HI is the baseline. Robust standard errors in parentheses clustered at the session level are reported in the parentheses. *** p<0.01, ** p<0.05, * p<0.1.

H2: Random individual effects probit regression analysis of product purchase decisions (IA vs. Control).

Independent variables	Dependent variable: $Buy_{j,t}$ =1, if buyer *j* purchased product in round *t* =0, o.w.					
	(1)IA and Control	(2) Marg. Eff	(3) IA and Control	(4) Marg. Eff	(5) IA and Control	(6) Marg. Eff
β_1: IA	0.484*	0.134*	0.078	0.020	-0.500	-.115
	(0.257)	(0.074)	(0.344)	(0.089)	(0.428)	(0.094)
β_2: IA_HI						
β_3: Round			-0.160***	-0.041***	-0.164***	-0.038***
			(0.027)	(0.007)	(0.028)	(0.006)
β_4: IA*Round			0.078*	0.020*	-0.097	-0.023*
			(0.046)	(0.012)	(0.054)	(0.135)
β_5: IA_HI*Round						
β_6: Advice *N*					1.26***	0.290***
					(0.346)	(0.067)
β_7: Advice *N* *Round					0.128***	0.029***
					(0.031)	(0.007)
Constant	0.076		0.951***		0.980***	
	(0.246)		(0.301)		(0.313)	
N	1120	1120	1120	1120	1120	1120

Note: Advice *N*=1 if Advice is N, and 0 o.w. Robust standard errors clustered at the session level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

H3: Random individual effects probit regression analysis of product purchase decisions (IA_HI vs. Control).

Independent variables	Dependent variable: Buy$_{j,t}$ =1, if buyer *j* purchased product in round *t* =0, o.w.					
	(1) IA_HI and Control	(2) Marg. Eff	(3)IA_HI and Control	(4) Marg. Eff	(5) IA_HI and Control	(6) Marg. Eff.
β_1: IA						
β_2: IA_HI	0.534	0.137	0.085	0.020	-0.084	-0.019
	(0.348)	(0.091)	(0.390)	(0.094)	(0.0371)	(0.084)
β_3: Round			-0.164***	-0.039***	-0.167***	-0.038***
			(0.028)	(0.007)	(0.028)	(0.007)
β_4: IA*Round						
β_5: IA_HI*Round			0.087**	0.021**	0.005	0.001
			(0.039)	(0.009)	(0.049)	(0.011)
β_6: Advice *N*					0.356	0.081
					(0.284)	(0.064)
β_7: Advice *N* *Round					0.100	0.024
					(0.323)	(0.016)
Constant	0.088		0.988***		1.00***	
	(0.259)		(0.317)		(0.323)	
N	1080	1080	1080	1080	1080	1080

Note: Advice *N*=1 if Advice is N, and 0 o.w. Robust standard errors clustered at the session level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

H4: Random individual effects probit regression analysis of shipping decisions (IA vs. Control).

Independent variables	Dependent variable: $Ship_{i,t}$ =1, if the seller i shipped the product in round t =0, o.w					
	(1) IA and Control	(2) Marg. Eff	(3) IA and Control	(4) Marg. Eff	(5) IA (advice N) and Control	(6) Marg. Eff
β_1: IA	0.869***	0.180***	-0.116	-0.024	0.531	0.096
	(0.282)	(0.047)	(0.360)	(0.077)	(0.353)	(0.059)
β_2: Noinsure					0.219	0.040
					(0.261)	(0.046)
β_3: Noinsure*IA					0.714**	0.129**
					(0.296)	(0.053)
β_4: Advice *N*			1.11***	0.231***		
			(0.267)	(0.064)		
Constant	-0.332*		-0.326*		-0.451*	
	(0.182)		(0.180)		(0.269)	
N	664	664	664	664	625	625

Note: Noinsure =1 if the buyer did not purchase the insurance; =0, o.w. Advice *N*=1 if Advice is N and 0 o.w. Robust standard errors clustered at the session level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

H5: Random individual effects probit regression analysis of shipping decisions (IA_HI vs. Control).

Independent variables	Dependent variable: Ship$_{i,t}$ =1, if the seller i shipped the product in round t =0, o.w					
	(1) IA_HI and Control	(2) Marg. Eff	(3) IA_HI and Control	(4) Marg. Eff	(5) IA_HI (advice N) and Control	(6) Marg. Eff
β_1: IA_HI	0.479	0.101	-0.130	-0.028	0.788*	0.153**
	(0.317)	(0.063)	(0.344)	(0.076)	(0.428)	(0.076)
β_2: Noinsure					0.216	0.042
					(0.257)	(0.049)
β_3: Noinsure*IA_HI					-0.043	-0.008
					(0.290)	(0.056)
β_4: Advice *N*			0.812**	0.177**		
			(0.027)	(0.058)		
			(0.267)			
Constant	-0.332*		-0.320*		-0.439*	
	(0.182)		(0.177)		(0.263)	
N	642	642	642	642	558	558

Note: Noinsure =1 if the buyer did not purchase the insurance; =0, o.w. Advice *N*=1 if Advice is N and 0 o.w. Robust standard errors clustered at the session level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Chapter 2: The Effect of Compassion Fade on Altruistic Behavior: Experimental Evidence for a Guilt Mitigation Account

Ben Grodeck[i], Toby Handfield[ii], Matthew Kopec[iii]

Abstract: In this paper we investigate the phenomenon of compassion fade. Specifically, where people act less altruistically when told that there are other people in similar circumstances who cannot be helped. Using variations of the dictator game, our design allows us to explore both the determinants of compassion fade, and the mechanism by which it operates. In our first experiment, which uses a between-subject design (n=187), we find that adding a second recipient to a dictator game who always receives a payoff of zero, regardless of the decision-maker's choice, significantly increases selfish behavior. We follow up with a large-scale experiment on Amazon mTurk using a within-subject design (n= 711). We replicate the result from Experiment 1. However, we also find that compassion fade is sensitive to the level of "need" – of the additional unhelpable recipient. When they are not in a needy state, decision makers act significantly less selfishly, compared to when they are in a needy state. Finally, we present evidence that the mechanism of compassion fade is diminishing negative affect (e.g., guilt) that arises from selfish decisions, rather than through the diminishment of warm glow.

JEL codes: D91, C91, D64,

Keywords: pro-social; motivated reasoning; altruism; compassion fade; identifiable victim

Acknowledgements: The Authors thank Christine Exley, Birendra Rai, Klaus Abbink, Justin Bruner, Philip Trammell, Mattie Toma, as well as seminar and conference participants from ASFEE (Lyon) 2022, ANZWEE (Sydney) 2022, ESA (Asia), Motivated Beliefs Workshop (Monash) 2022, EA Global (San Francisco) 2019), ICSD (Arizona) 2019. TH's work on this project was supported by the Australian Research Council, grant FT180100067. This project was also funded by the Australian Research Council Discovery Early Career Researcher Award (DE180101119)

[i] Monash University, Department of Economics, Email: ben.grodeck@gmail.com (*corresponding author)
[ii] Monash University, Department of Philosophy, Email: toby.handfield@monash.edu
iii Northeastern University, Department of Philosophy, Email: m.kopec@northeastern.edu

1. Introduction

Humans have a substantial motivation to act altruistically towards others. However, a growing body of evidence suggests that contextual differences influence altruistic behavior in normatively questionable ways. There are well known violations of classical axioms of Expected Utility Theory, such as framing effects[22] (Tversky & Kahneman, 1981), hyperbolic discounting (Loewenstein & Thaler, 1989) and ambiguity aversion (Ellsberg, 1961) all of which have been found to affect altruistic decision-making. There are also various phenomena that seem to affect altruistic behavior more particularly. People engage in self-serving justifications of various sorts (Gino et al., 2016), including motivated reasoning (Benabou & Tirole, 2002), information avoidance (Capra & Larson, 2009; Dana et al., 2007; Grossman & der Weele, 2013; Matthey & Regner, 2011; Serra-Garcia & Szech, 2019), and self-serving interpretations of morality (Hamman et al., 2010; Shalvi et al., 2011) and of reality (Di Tella et al., 2015; C. Exley & Kessler, 2019; C. L. Exley, 2016a; Kassas & Palma, 2019).

A specific phenomenon that has attracted less attention in economics, but is well documented in psychology is "compassion fade", whereby altruistic behavior diminishes when more needy individuals are included in the decision frame, even if the choice set of the decision maker are the same (Butts et al., 2019; Erlandsson et al., 2014; Markowitz et al., 2013; Small et al., 2007; Västfjäll et al., 2014).[23] A typical compassion fade experiment has a control treatment in which a participant is asked to donate $20 to help a particular individual who faces adverse circumstances. The treatment adds mention that there are other people in similar circumstances that cannot be helped. These studies find that altruistic decisions diminish significantly in the treatment condition (Erlandsson et al., 2016; Markowitz et al., 2013; Slovic, 2007; Small et al., 2007; Västfjäll et al., 2014).[24]

---

[22] We take no stand on whether framing effects are strictly irrational (cf. Horne & Livengood, 2017).

[23] While compassion fade can also include problems of distributing money to many helpable recipients, this paper focusses on a thin notion of compassion fade where only one repiient can be helped.

[24] What we call compassion fade is evidently related to the "identifiable victim effect" whereby identifiable individuals elicit stronger altruistic responses than more anonymous, "statistical" individuals. However, compassion differs from the identifiable victim effect, since the latter refers to different behavior towards individual recipients on the basis of information about that specific recipient. Whereas we take compassion fade to be a diminution of altruistic behavior conditional on information about others who cannot be helped.

The basic intuition behind compassion fade is that we will be less motivated to help an individual when told that there are other people who are suffering or in need, but cannot be helped. This suggests that there may be two distinct factors that contribute to compassion fade: The first is the number of unhelpable individuals in the decision frame. Simply being aware that there are others who cannot be helped may reduce one's altruistic behavior. The second is the degree of need of those unhelpable individuals. It might not be relevant to a decision maker if the additional individuals are relatively well off, but altruistic behavior may decrease if the additional individuals are especially needy.

Although the phenomenon of compassion fade has been well documented, there are of the existing research. First, the two components of compassion fade have not been consistently distinguished. Second, the extent and magnitude of compassion fade has not been quantified. Further, it has been investigated primarily in the context of charitable giving: in a typical experiment, participants are asked to donate to address a publicly known instance of serious deprivation, and there is explicit involvement of a charitable institution in delivering the intended benefit. But compassion fade may be relevant for much more mundane instances of cooperative and pro-social behavior, such as within firms, voluntary organizations, or in random encounters with strangers. One of our experiments uses an anonymous interaction between mTurk workers – thus is relevant to the existence of compassion fade amongst peers (Almaatouq et al., 2020), and the mTurk platform provides a relatively direct and straightforward intermediary to effect the transaction, thus avoiding any effects that may be idiosyncratic to the involvement of charitable organizations. Finally, compassion fade has been investigated typically using between-subject experiments, (with the exception of (Bartels, 2006) and (Västfjäll et al., 2014), which may lead to effects which would be substantially moderated in a within-subject experiment. This is because participants can more readily detect inconsistencies in their choices when faced with multiple decisions in a within-subject design.

These limitations in the existing literature give rise to the following research questions. Does compassion fade exist in non-charitable giving environments? Can we distinguish the effect of the different dimensions that may contribute to compassion fade: the degree of need and the number of needy individuals in the decision frame? Third, are these effects robust when people make multiple decisions and have the opportunity to notice possible inconsistency across decisions? Finally, what is the mechanism by which compassion fade operates: is it due to diminished disutility ("guilt") for selfish decisions, or diminished warm glow for generous decisions?

We conducted two experiments, using variations on the classic dictator game, with either one or two recipients. The decision maker chooses between two alternative distributions of money for themself and the recipients. The first experiment was a traditional lab experiment, using a between-subject design. We found that adding a second recipient who always got a payoff of zero more than doubled the rate of selfish behavior. This provides evidence that compassion fade plays a role in 3-player dictator game environments. This contributes to the literature on dictator games with multiple recipients. Although dictator games are commonly used in economics, the behavior of dictators when the framing and structure of these games changes from one to two potential recipients is still poorly understood (Bolton et al., 1998; Bolton & Ockenfels, 2008; Fisman et al., 2007)

Given many framing effects show much smaller effect sizes when studied within subjects (Kahneman & Tversky, 1996), we thought it important to conduct a within-subject experiment which would enable subjects to try to be consistent with classical decision norms,[25] thereby minimizing compassion fade.

Our second experiment was a large-scale study conducted on Amazon mTurk, using a within-subject design. We designed this experiment to empirically distinguish two dimensions of compassion fade. The design also allows us to draw some indirect inferences about the mechanism of compassion fade. Our results are broadly consistent with the findings of Experiment 1, though with a reduced magnitude of effect.

We observe an 11.3% increase in selfish behavior when a second recipient—who receives a payoff of 0 no matter what—is added to the dictator game environment. However, we find that when this second, unhelpable recipient is not in a needy state (payoff of 10), dictators are significantly less likely to act selfishly compared to when the unhelpable recipient is in a needy state (payoff of 0). This shows that the unhelpable recipient's level of neediness influences altruistic decisions.

Using our repeated measurements of participants in Experiment 2, we constructed a (pre-registered) typology of subjects based on different choices they made in the decision tasks. Our 7-category typology is able to categorize 77.4% of participants. Of the categorizable participants, we find 12.2% have compassion-faded preferences. Overall, our results are consistent with the findings from Experiment 1. While the magnitudes are not as

[25] Nielsen and Rehbeck (2022) provide evidence that participants in experiments have a preference of acting consistently with expected utility axioms.

large, our results suggest that there are a substantial group of participants who consistently respond to these different framing effects.

Given that compassion fade is sensitive to the decision maker's information environment, it remains an open question whether decision makers are aware of this sensitivity, and if so whether they would manipulate their information environment when given ambiguous information. As we explain in section 2, different hypothesized mechanisms underlying compassion fade make different predictions as to whether or not participants would prefer to experience compassion fade or not. In a variation on the moral wiggle room experiment (Dana et al., 2007) we observed that indeed, participants will avoid information about the payoff to an unhelpable third party, and disproportionately make the selfish choice in that case. This provides indirect evidence that participants prefer to believe that they are in a context which would prompt compassion fade; this suggests compassion fade is a subjectively "desirable" phenomenon, which in some way diminishes negative utility associated with a selfish decision, rather than diminishing a possible "warm glow" associated with an altruistic decision.

The rest of our paper proceeds as follows: Section 2 describes the motivating framework. Section 3 explains the experimental design, hypotheses, and results of experiment 1. Section 4 discusses the experimental design, hypotheses, and results of experiment 2. Section 5 discusses the implications of our findings. Finally, Section 6 summarizes the main findings of the paper.

2. Motivating Framework

In this section we elaborate the mechanism in which compassion fade might operate. First, we distinguish two components of compassion fade: the number of unhelpable recipients in the decision frame, and the level of wellbeing of these unhelpable individuals. Then we discuss whether compassion fade operates by reducing positive utility associated with being altruistic, or mitigating negative utility associated with being selfish.

Regarding the two components of compassion fade, see Table 1 for a series of cases that might be used in a typical compassion fade experiment to distinguish these two components. In these vignettes the decision task at hand is whether you should help Rokia or not.

Table 1. Compassion Fade Phenomena

	Vignette	**Compassion fade phenomena**
Control	You can help Rokia, a malnourished child from Mali, with your donation of $20.	n/a
A	You can help Rokia, a malnourished child from Mali, with your donation of $20. **There are thousands of other children in Mali who are malnourished.**	Compared to a control, which omitted the last sentence, vignette A will elicit less altruism. But it is unclear whether this is due to the number of children mentioned in the second sentence, or the degree of their suffering.
B	You can help Rokia, a malnourished child from Mali, with your donation of $20. There are **tens of** thousands of other children in Mali who are malnourished.	If number of needy individuals who cannot be helped is relevant, we expect less giving in B than A.
C	You can help Rokia, a malnourished child from Mali, with your donation of $20. There are thousands of other children in Mali who are **severely** malnourished.	If the degree of need of the unhelpable individuals in relevant to compassion fade, we expect less giving in C than A.[26]

We suggest compassion fade can be understood as a function of a decision problem akin to a dictator game with a number of additional, unhelpable individuals who resemble the recipient and are part of the context of decision. The prediction of compassion fade is that generous behavior will diminish as the number of unhelpable individuals increases and as the wellbeing of the unhelpable individuals decreases. This is notwithstanding that the choice faced by an individual is the same, from a classical rational choice perspective (Loomes, 1991; Machina, 1987).[27]

Figure 1Figure 1 illustrates the preferences of the dictator. The horizontal and vertical axes represent the dictator and the helpable recipient's payoffs (x_i and x_j) respectively. Line AB is an exogenously fixed budget line.

Figure 1(a) shows what happens when we introduce an unhelpable recipient to the decision frame. We assume that adding this unhelpable recipient to the dictator game changes the dictator's preferences such that they prefer more selfish allocations. As a result, the

[26] One might be sceptical that the word "severely" would be enough to generate this compassion fade phenomena. Our experiment uses quantitative differences in the level of need of the unhelpable recipient for this reason.

[27] It may be possible to further generalize this definition to include cases where the choice set itself changes, to allow the decision maker to help more people. But it is difficult then to be sure that any change in generosity is due to a drop in compassion or is instead due to changes in the marginal cost of following a consistent policy.

indifference curve of the dictator becomes more vertical: shifting from a to a'. Since we have the same budget line (AB), dictators will choose more selfish allocations.

Figure 1(b) shows what we conjecture happens when we then increase the payoff of the unhelpable recipient: the effect of compassion fade is to some extent mitigated. Consequently, the indifference curve of the dictator shifts from a' to a''. As a result, dictators will choose less selfish allocations when the payoff of the unhelpable recipient increases.

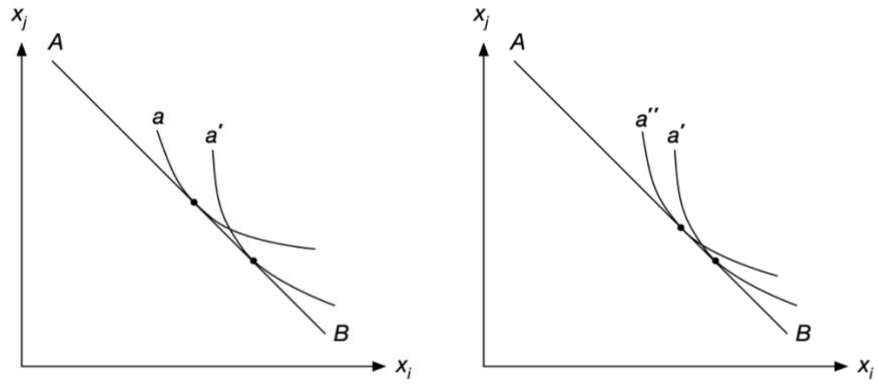


Figure 1. a: The indifference curves of the dictator when unhelpable recipients are added (a'). b: The indifference curves of the dictator when the payoff of the unhelpable recipients increase (a'').

In sum, we make the following initial assumptions about compassion fade.

- As the number of unhelpable recipients increases, compassion fade increases
- As the payoff of the unhelpable recipient increases, compassion fade decreases
- When there are no unhelpable recipients, there is no compassion fade
- When there are unhelpable recipients who receive nothing, then compassion fade has a non-zero effect

Moving beyond these highly schematic claims about the effect of compassion fade, if we were to move to a fully specified model, we would need to choose whether compassion fade increases selfish behavior (Figure 1a) by either reducing the magnitude of a negative utility term associated with selfishness (which might represent guilt, shame, or similar psychological phenomena), or by reducing the magnitude of a positive utility term associated with generosity (which might represent warm glow or similar).

To illustrate, compare the utility functions described below where x_i is the decision maker's playoff, x_j is the helpable recipient's payoff, and x_k is the unhelpable recipient's payoff. The decision maker's choice (how much to give) only affects their payoff x_i and the helpable recipient's payoff x_j. In both utility functions compassion fade $c(x_k)$ moderates the magnitude of the utility associated with decision's impact on the payoff of the helpable recipient. In model (1) $g(x_j)$ represents the disutility associated with ungenerous allocations,[28] and in model (2) $h(x_j)$ represents a positive utility component associated with generous allocations (such as warm glow):

$$(1)\ u(x_{i'}, x_j, x_k) = x_i - (1 - c(x_k)) * g(x_j)$$
$$(2)\ u(x_{i'}, x_j, x_k) = x_i + \left(1 - c(x_k)\right) * h(x_j)$$

In general, model 1 suggests that if they had the opportunity, a selfish individual would prefer to induce more compassion fade rather than less. Whereas model 2 suggests that participants would minimize compassion fade if they could, because compassion fade leads them to miss out on additional utility from warm glow.

In explaining model 1 above, we refer to "disutility". This term has the advantage of being highly general and non-committal: it may refer to guilt, understood as negative affect associated with failure to meet the expectations of others (cf. Battigalli & Dufwenberg, 2007); or shame, understood as negative affect associated with loss of esteem or the violation of social norms (Sznycer et al., 2018); or simply to an aversion to violating a personal norm (Capraro & Perc, 2021). For convenience, we will call accounts of any of these types "guilt-mitigation" accounts, but it should be understood that we are not ruling out any of these alternative approaches.

3. Experiment 1

We originally designed Experiment 1 to investigate a different phenomenon.[29] However, in this section we only focus on the compassion fade hypothesis in this experiment.

3.1. Treatments

Figure 2 shows the control treatment—a binary dictator game with 1 recipient who can be benefited by a lesser or a greater amount, at a modest cost to the dictator. The dictator can choose either Option 1 where they receive \$10 AUD and the recipient receives \$10, or Option 2 where they receive \$12, and the recipient receives \$2. In all treatments we told

---

[28] Saccardo and Serra-Garcia (2020) investigate a different sort of decision problem, with the feature that participants can choose their information environment. They find that participants are more likely to choose an environment in which they can justify less altruistic behavior, rather than more altruistic. This suggests a similar picture to model 1.

[29] See Appendix A for more background on the original experiment and the other treatments that were part of the experimental design.

participants that Option 1 was the original allocation, making the generous allocation the implied status quo.[30]

Figure 3 shows the Three-Player Game treatment (3PG(0) from here on). This treatment has the same structure as the control, except we add a second dummy recipient to the framework. The dictator is unable to affect the payoff of this dummy recipient, who receives $0 no matter what option the dictator chooses.

	Dictator	Recipient 1
Option 1	10	10
Option 2	12	2

Figure 2. Control

	Dictator	Recipient 1	Recipient 2
Option 1	10	10	0
Option 2	12	2	0

Figure 3. 3PG(0) treatment

3.2. Procedure

The experiment was conducted at Monash University in the Monash Laboratory for Experimental Economics (MONLEE), using oTree software (Chen et al., 2016). Our decision-makers were recruited from the MONLEE participant pool. Upon arrival, Participants were randomly allocated their seats. The experimenter then read out the instructions for each task. To ensure understanding of the main task, participants answered simple comprehension questions. Then they made their decision. Upon finishing the main task, participants also completed The Oxford Utilitarian Scale (Kahane et al., 2018),[31] and an

[30] These values represent the same ratios for the dictator and recipient as in (Dana et al., 2007).

[31] This measures how utilitarian someone is on two scales. Impartial Beneficence, the idea that helping others should be from an impartial perspective. Instrumental Harm, the idea that harming someone to make others better off is acceptable. We do not find any interesting correlations and thus do not include it in our analyses.

indirect measure of subject's credence (in aggregate) that we will pay the recipients as prescribed by the instructions (Glynn, 2013), and demographic questions.

Recipients were recruited from a pool of students in large undergraduate history and philosophy classes on campus. Prior to the laboratory session, these recipients were either randomly allocated to the role of recipient in the control treatment, or recipient A or B in the 3PG(0) treatment with equal probability. After the laboratory session was over, recipients were invited via email to attend a short session on campus explaining the experiment, what role they were allocated to, and how their payoff was determined. Recipients were then paid in cash.

We had 187 participants complete this study: 110 in the control treatment and 77 in the 3PG(0) treatment, across 9 different sessions.[32] Each session lasted less than 45 minutes. There was no show up fee.[33] On average, participants earned $10.8 AUD, while recipients on average earned $4.9 AUD.[34] The experimental design, predictions, power analyses and procedure were pre-registered on https://osf.io/uwyqp/.

3.3. Hypotheses

We have one dependent variable of interest: whether the participant chooses Option 1 (generous) or Option 2 (Selfish). Our null hypothesis is that there is no difference in rates of selfish behavior between the control and 3PG(0) treatments.

If we find a significantly higher rate of selfishness in 3PG(0), this suggests the addition of a third party to the scenario matters: even though the decision cannot affect the payoff of this third party, it changes the relative attractiveness of the two options. This is what models of compassion fade would predict. This leads us to the following prediction:

H1: The proportion of selfish choices will be higher in the 3PG(0) treatment compared to the control treatment.

3.4. Results

Figure 4 presents the proportion of participants choosing selfishly (Option 2) in each treatment. We observe that 26.4% of participants make the selfish choice in Control. This

[32] Since the 3PG(0) treatment was originally used as a control for Study 1 we recruited less participants in this treatment. While this is a limitation in the interpretation of the results of Study 1, the results from this treatment motivated us to follow up with Study 2.
[33] Since Participants earnt at least $10 AUD in this experiment, we decided to have no show up fee.
[34] This includes the 77 recipient Bs in the 3PG(0) treatment who received $0.

finding is consistent with the results from Dana et al., (2007) which makes sense given that our payoff structure in the control has the same ratios as their study. However, compared to the Control treatment, we find a significant increase in the rate of selfish behavior in 3PG(0): 55.8% vs 26.4% ($\chi^2 = 16.6$, $p < 0.001$).

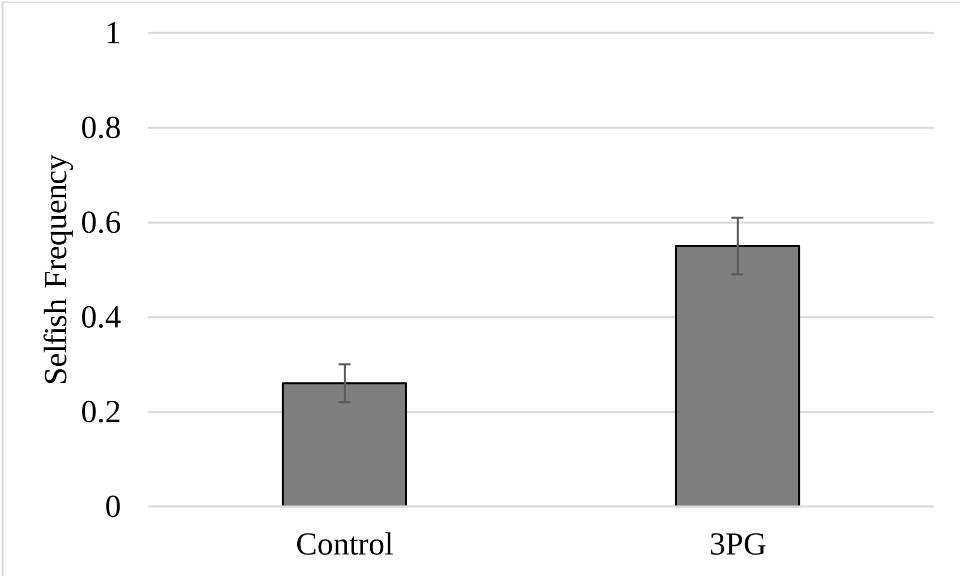


Figure 4. Frequency of Selfish Decisions by Treatment in Experiment 1.

We also find that participants in both treatments found our experimental task to be credible, in the sense that they believed that we would pay the recipients as described in the instructions. To assess our credibility, we used a list randomization method (Glynn, 2013), rather than a direct method which is likely to generate substantial experimenter demand effects (Zizzo, 2010). In this method, participants are randomized into two groups. One group received four dummy statements[35] and the other group received the same four statements with one extra: "*the experimenter will pay the recipients as described in the instructions*." Participants were then asked to report how many statements they believed to be true. Since only one group had the statement of interest, we can indirectly estimate the proportion of participants who held this belief.

Despite our unorthodox design,[36] in which each decision maker was promised that their decisions would affect payments to individuals who were not present in the lab, we observe that participants found the instructions to be believable. We estimate that the mean credence in our proposed payment procedure was 0.86 (95% CI, 0.66–1.00).

[35] See Appendix B for Experiment 1 instructions

[36] This design choice was made as one of the original treatments, some recipients may not learn that they were even part of the experiment. See Appendix A for more details.

3.5. Discussion

Our findings show that the addition of an “unhelpable” second recipient to a Dictator game affects the behavior of dictators, substantially increasing the frequency of selfish behavior. The difference in rates of selfish behavior hints at a possible violation of the axiom of independence of irrelevant alternatives. However, as this experiment used a between subject design, we cannot say with certainty that participants violate this axiom, since they only made a single choice.

4. Experiment 2

While compassion fade and its components have been well documented in charitable giving settings, our first experiment shows that compassion fade also arises in a different setting: where the decision maker and the recipient are peers; no charitable organization is mediating the transaction; and where there is no exceptionally high level of neediness for the recipient. (Although the typical participant in our lab is not wealthy, they are much less needy than the recipients in typical charity-based experiments.)

Our first experiment leaves several questions about the mechanism of compassion fade unanswered. In particular, although we have theorized that there are two dimensions which may contribute to compassion fade – the number of unhelpable individuals and the degree of their neediness – it remains untested whether both of these factors make a detectable contribution to the phenomenon.

Second, we have canvassed that there are a number of possible mechanisms by which compassion fade may work: broadly speaking, a compassion faded individual may lose positive utility normally associated with generous behavior (e.g., warm glow), or they may mitigate negative utility normally associated with selfish behavior (e.g., guilt or shame). Our second experiment includes a treatment designed to differentially confirm these two possibilities, by putting participants in an ambiguous setting where they can choose to remain ignorant as to the exact decision problem they face. The intuition behind this design is that some individuals may be able to deceive themselves so as to increase their credence in a utility-maximizing state. Individuals who are not subject to such credence manipulation have a reason to reveal the underlying game prior to their decision. But any trend in the decisions by individuals who choose to remain ignorant allows us to infer the structure of those individuals’ underlying utility functions.

If participants lose warm glow by compassion fade, they should seek to convince themselves that the unhelpable recipient is well off, which will prompt them to behave more generously and obtain the warm glow of giving. On the other hand, if compassion fade causes a diminishment of guilt, then self-deceiving participants should instead be motivated to convince themselves that the unhelpable recipient is poorly off, which will enable them to not donate without experiencing the cost of guilt.[37]

Finally, there are other possible explanations for the increase in selfish behavior other than compassion fade, and we designed our second experiment to rule some of these out. Putting aside payoffs to the decision maker, in our 3PG(0) treatment from experiment 1, Option 1 generates more inequality between the recipients. If subjects are averse to creating such inequalities, this might contribute to a preference for Option 2. Hence, we have two "NoProfit" treatments which test whether, in the absence of any motive of personal gain, participants have any preference to minimize such inequalities.

In addition to these questions about the mechanism of compassion fade, in Experiment 2 we use a within-subjects design to assess the magnitude and persistence of the phenomenon. We developed a pre-registered typology of possible preference structures that participants might exhibit, and report proportions of participants who exhibit choices consistent with those types.

4.1. Treatments

An overview of our treatments of interest is presented in Figure 5.

[37] There is already some evidence that compassion fade is a phenomenon that is sensitive to considerations of self-interest: Cameron & Payne (2011) find that participants anticipate experiencing compassion fade, understood psychologically, as a diminishment of empathy or compassion, only when they expect to be later asked to make a donation to alleviate the problem.

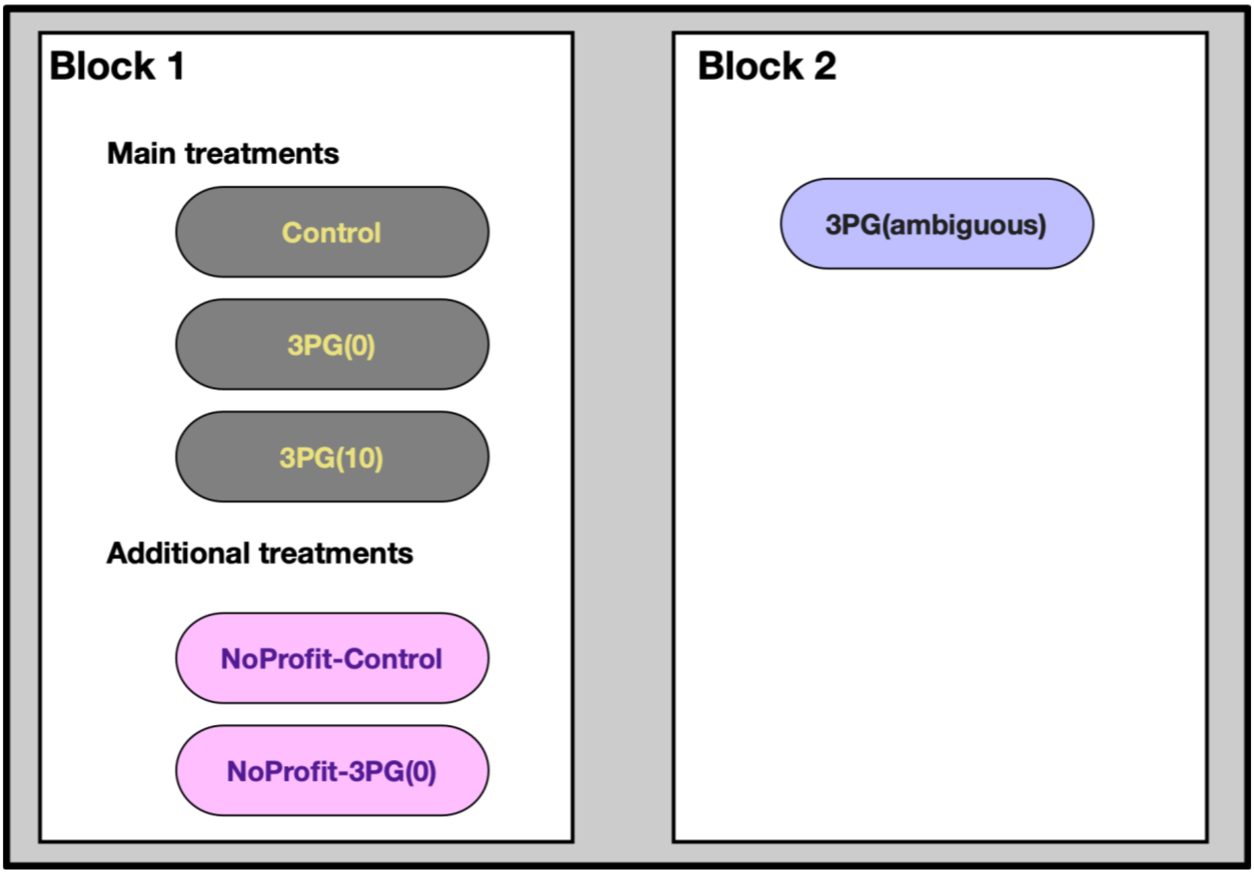


Figure 5. Overview of Main treatments in Experiment 2. Omitted from this diagram for clarity are three additional treatments which were not relevant to the hypotheses of this paper. See Appendix C for full details.

Table 2. Treatments from Block 1. The final column indicates the payoffs associated with the two options in each treatment.

Category	Label	Is helping costly?	Unhelpable individuals?	Neediness of unhelpable	Options Payoffs to (Donor, Recipients)
Main					
	Control	Yes	No	—	(10,10) (12,2)
	3PG(0)	Yes	Yes	High	(10,10,0) (12,2,0)
	3PG(10)	Yes	Yes	Low	(10,10,10) (12,2,10)
Additional					
	NoProfit-Control	No	No	—	(10,10) (10,2)
	NoProfit-3PG(0)	No	Yes	High	(10,10,0) (10,2,0)

We summarize the treatments from Block 1 in Table 2. **Control** is the two-player control game from Experiment 1. **3PG(0)** is the same as our 3PG treatment from Experiment 1: an additional "recipient" is added to the decision frame, but is guaranteed to receive zero, no matter what option the decision maker takes. **3PG(10)** differs from **3PG(0)** treatment in that the second recipient will now receive 10 ECU, regardless of the decision maker's choice. The comparison of **3PG(0)** and **3PG(10)**, therefore, allows us to assess what difference the relative neediness of the additional recipients makes to compassion fade.

In addition to these three main treatments, we had three additional treatments. The two **NoProfit** treatments – **NoProfit-Control** and **NoProfit-3PG(0)** – are just like our Control and 3PG(0) treatments, except that we have removed the selfish incentive for the decision maker not to help. The decision-maker's payoff is fixed at 10. This allows us to assess whether the decision maker has distributional preferences between the two recipients that might lead to less generosity than we would otherwise expect.

Comparing the five treatments from Block 1 permits us to systematically compare the decisions across games that only differ in a single variable. We are able to compare the following.

- Comparing Control and 3PG(0) allows us to investigate the impact of adding an unhelpable recipient who receives nothing.
- Comparing 3PG(0) and 3PG(10) allows us to investigate the impact of the payoff of the unhelpable recipient.
- The difference between Control and 3PG(0) compared to the difference between NoProfit-Control and No-Profit(3PG(0), allows us to investigate the impact of varying the degree of inequality between the recipients.[38] (3PG0-Control) – (NoProfit3PG0-NoProfitControl).

To control for order effects, we randomize the order in which each task is presented to the participants. After completing Block 1, participants then entered Block 2, where they undertake the 3PG(ambiguous) treatment.

The 3PG(ambiguous) treatment is inspired by recent experiments using the so-called "Moral Wiggle Room" game (Dana et al., 2007). In that paradigm, the decision maker has incomplete information about the impact of their decision on others, but has the opportunity to obtain complete information before deciding. Some participants in that design evidently exploit their ignorance as an opportunity to engage in motivated reasoning – making decisions that are inconsistent with their behavior under full information. Such designs thus allow us to draw inferences about motivations that may not be readily apparent otherwise. In our case, participants are either in the 3PG(0) treatment or the 3PG(10) treatment. Decision-makers thus have full knowledge of their own payoff and the payoff of Recipient 1. However, they do not know whether recipient 2 will be receiving 0 or 10. Each state has a 50% chance of occurring. Participants can choose to reveal (at no cost) which possible world they are in, or choose either Option 1 or Option 2 in ignorance.

Using this game, we are interested in focusing on participants who exhibit sensitivity to the neediness dimension of compassion fade: that is those who chose generously in 3PG(10) but selfishly in 3PG(0). For these participants we can distinguish the following types:

- Consistent types: choose to reveal, and then play the same way as they did in the equivalent full information game.
- Perverse types: choose to reveal, and then do not conform to how they played in the equivalent full information game.

[38] If the decision maker cares about the inequality between the recipients' payoffs, they may be choosing the selfish option in 3PG0 to minimise this inequality. By removing the selfish incentive in the two No-Profit conditions, we can test whether any increase in the selfish option is due to the decision maker wanting to minimise inequalities between the recipients.

- Fade biased types: choose not to reveal, and choose selfishly
- Compassion biased types: choose not to reveal, and choose altruistically.

As the name implies, the only type that we would expect using a "classical", rational representation of this decision task would be the consistent type — if there is a utility benefit to choosing differently in the two different scenarios, and the information is free, the information should be obtained. But if we observe a substantial proportion of participants falling into one of the other three types this suggests a possible deviation from standards of rationality, and may shed light on the mechanism by which compassion fade occurs.

Inspired by the recent literature on motivated reasoning and self-image (Benabou & Tirole, 2002; C. Exley & Kessler, 2019; C. L. Exley, 2016b, 2016a; Grossman & der Weele, 2013; Serra-Garcia & Szech, 2019), we claim the two utility functions from section 2 will give rise to different predictions when the dictator encounters an environment of information uncertainty.

If participants can manipulate their beliefs under uncertainty, then utility function (1) predicts that compassion faded types will, in 3PG(ambiguous), not reveal the payoff of the unhelpable recipient and choose selfishly. This is because compassion fade reduces a negative utility term (guilt or similar), so the agent will prefer to believe that the unhelpable individual is poorly off, and thus can experience less guilt.

However, in similar conditions of incomplete information, utility function (2) predicts that participants would act more pro-socially. As compassion fade reduces the positive utility term (h), we would expect "self-serving" beliefs to promote the belief that the unhelpable individual is well off, thus making more altruistic behavior in 3PG(ambiguous) the utility-maximizing decision.

As pre-registered, we predicted that the guilt mitigation style of account in (1) would be more likely to occur. Consequently, the hypotheses below are based on that account.

4.2. Hypotheses

Using the results from Experiment 1 and the motivating framework from section 2, we generate the following hypotheses.

In all cases, our principal measure is the rate of selfish behavior (choosing Option 2) in each game.

Hypothesis 1: The rate of selfish behavior will be greater in 3PG(0) than in Control. This is a replication of Experiment 1, where we predict that adding a unhelpable recipient who gets 0 no matter what, will result in an increase in selfish behavior due to compassion fade.

Hypothesis 2: The rate of Selfish behavior will be greater in 3PG(0) than in 3PG(10). Since we predict that the less needy the unhelpable recipient is, the less selfish the decision maker will be as compassion fade decreases.

Hypothesis 3: We predict that participants will use the lack of information in the decision environment to act more selfishly in 3PG(ambiguous). More precisely, we distinguish between two types.

- Non discriminating types: choose the same way in 3PG(0) and 3PG(10).
- Discriminating types: choose differently in 3PG(0) and 3PG(10).

Our null hypothesis is that both these types of participants are consistent with their choices in 3PG(ambiguous). Put in positive terms, we hypothesize the disjunction of the following three claims:

H3a (Non discriminating type): The rate of selfishness in 3PG(ambiguous) will differ from the average rate of selfishness in 3PG(0) and 3PG(10).

H3b (Discriminating types that choose to reveal): The rate of selfishness in 3PG(ambiguous), given that the participant reveals the true state in this game, will differ from the rate of selfishness in the equivalent version of the game (3PG(0) or 3PG(10)).

H3c (Discriminating types that choose not to reveal): The rate of selfishness in 3PG(ambiguous), given the participant does not reveal the true state of the world will not equal 50%. (Given that in one of 3PG(0) and 3PG(10) they choose Option 2, and in the other they choose Option 1, if they are choosing randomly, we should not expect them to be more likely to choose one option over the other.)

4.3. Procedure

The experiment was conducted on Amazon Mturk using Cloud Research (Litman et al., 2017).[39] To ensure understanding, participants were required to answer a number of comprehension questions before completing each of the nine tasks (See Appendix D for the full set of instructions and comprehension questions). We told participants that one of the nine tasks would be randomly chosen for payment at the end of the experiment. Participants were recruited from the United States[40] and had to have an approval rate of at least 85% in previous Amazon Mechanical Turk HITs to be eligible for the study.

Before recruiting participants, we invited a group of recipients to act as the helpable and unhelpable recipients.[41] Participants were told that these recipients had already been recruited and that they would be matched with either one or two of these recipients (depending on which task was chosen for payment). Participants were explicitly told that they would be uniquely matched with these recipient(s) in a predetermined manner, so that their choice is the only one that would affect the recipient's payoff.[42] We informed the participants that we would explain to the recipients what the randomly chosen task was, and how their payoff was determined as a result of the decision maker's choice.

After completing all the decision tasks, participants also completed the indirect measure of subject's credence (in aggregate) that we will pay the recipients as prescribed by the instructions, and some demographic questions. To be consistent with experiment 1, we kept the same numbers and ratios in the games, however we used Experimental Currency Units (ECUS) where 1 ECU = $0.2 USD.

We had n=852 participants complete the study.[43] The average time taken to complete the experiment was just over 15 minutes. On average, participants earned $2.15. Those in the

---

[39] The use of platforms like MTurk (or Prolific) has already received significant uptake in experimental/behavioral economics (C. Exley & Kessler, 2019; C. L. Exley, 2019; Gandullia, 2019; Gandullia et al., 2020; Gandullia & Lezzi, 2018; Hauser & Schwarz, 2016; Palan & Schitter, 2018; Serra-Garcia & Szech, 2019). As Gandullia et al., (2020, p. 2) have argued, moving from a university student sample to an online sample may also reduce experimenter demand effects as the experimenters are not physically present at the time of data collection thus further making plausible this choice of participant recruitment

[40] Even though we use low stakes in Experiment 2, Raihani et al. (2013) find no difference in behavior from US Mturk participants when stakes were low compared to high.

[41] Recipients were excluded from being participants.

[42] See instructions in Appendix D

roles of recipient A earned \$1.39 and recipient B earned \$0.60. The experimental design, predictions, power analyses and procedure were pre-registered on https://osf.io/uwyqp/.

4.4. Results

In accordance with our preregistered plans, we excluded participants whose comprehension error rate was more than 1 standard deviation greater than the mean. Of the remaining 711 participants,[44] the mean error rate for comprehension questions was 3.64 (out of 35 comprehension questions).[45]

First, we look at the rate of selfish behavior in the tasks from Block 1 (see Figure 6). We find that the rate of selfishness is 11.3% higher in 3PG(0) compared to Control (35.3% vs. 31.7%, McNemar test, p=0.028), confirming **H1**. This result suggests that adding an unhelpable person who is "needy" (payoff of 0) to the decision environment increases the rate of selfish behavior. After adjusting for multiple hypothesis testing—see details in Appendix E—we find that the probability of finding this result if the data came from the same distribution as the Control is slightly higher than our nominated alpha level of 0.05, but would still be considered statistically significant at a critical value p = 0.1. This result provides modest evidence in favor of **H1** that the rate of selfish behavior increases when an unhelpable recipient is added to the decision environment.

We test whether the effect observed here could be entirely due to the decision maker being averse to inequality between recipients. If participants are more likely to choose the inequality-minimizing allocation (the "ungenerous" one) when there is no personal profit motive, this would provide evidence that inequality concerns are motivating their decisions. We find no evidence of this. The rate of choosing the inequality-minimizing allocation does not significantly differ between the conditions Noprofit-Control and NoProfit-3PG(0) (9.6% vs. 11.5%, McNemar test, p=0.184).

---

[44] As written in our pre-registration, we aimed to have n=720 participants after excluding those who with high comprehension errors. Given there are 6!=720 possible orders of the first block of decision tasks, this was to ensure the orders were counterbalanced.

[45] Similarly to Experiment 1 we find that participants in both treatments found our experimental task to be credible, in the sense that they believed that we would pay the recipients as described in the instructions. Using the same indirect elicitation method, we estimate that the mean credence that we would adhere to our described payment procedures is 0.82 (95% CI 0.65–1.00). This indicates that subjects had a high credence in the incentive compatibility of their decisions.

Further, we test within subjects whether they are more likely to exhibit the sort of inconsistency typical of compassion fade across the profit and no-profit conditions. Choosing to give generously in Control and ungenerously in 3PG(0) could be due to a concern to reduce the inequality between the two recipients in 3PG(0) (helpable and unhelpable). This motivation, however, would predict a similar pattern in the NoProfit-Control and Noprofit-3PG(0) tasks. To test this, we create two new dummy variables, where we code a decision as 1 if the participant chooses Option 2 in 3PG(0) and Option 1 in Control, and as 0 otherwise. We do this for both the profit and NoProfit cases respectively. This resembles a test of difference-in-differences.

(3PG0 – Control) – (NoProfit-3PG0 – NoProfit-Control)

We test the null hypothesis that the probability of this pattern of decision is the same within subjects. We can reject this null hypothesis. We find significantly more selfish behaviour occurs in the profit conditions than in the NoProfit conditions (11.7% vs. 7.7%, McNemar test, p=0.010). These results suggest that the increase in selfish behavior when there is a profit incentive is at least partly due to compassion fade. While we do not find evidence of inequality aversion, we do not rule out that there may be some of these inequality effects at play.

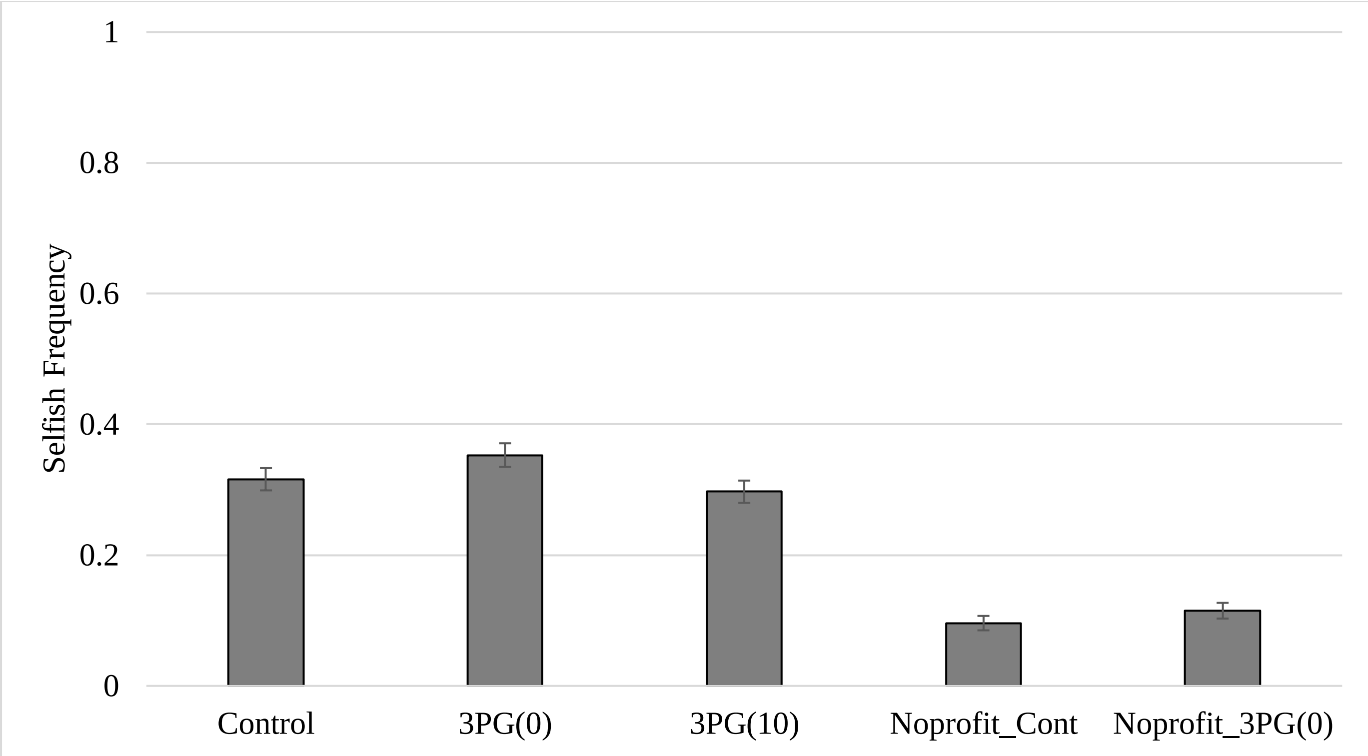


Figure 6. Frequency of Selfish Decisions by Treatment in Experiment 2.

Next, we test **H2**. As predicted, we find that the rate of selfish behavior is significantly higher in 3PG(0) than 3PG(10) (35.3% vs. 29.7%, McNemar test, $p < 0.001$). We also do not find a difference between 3PG(10) and Control (M29.7% vs. 31.7%, McNemar test, p=0.268). However, by testing for mean equivalence (Schuirmann, 1987; Dinno, 2017), we find evidence for this statistical equivalence between 3PG(10) and Control. This null result is bounded by 5 percentage points in either direction (lower bound, p=0.033; upper bound, p<0.001). If we only consider at the null hypothesis that selfish behavior is higher in 3PG(10) than the Control by more than 1 percentage point, we can reject this null hypothesis (p=0.036) showing the upper bound of any potential increase in selfishness in 3PG(10). These results suggests that the unhelpable recipient's payoff is an important component of compassion fade and mitigates the original effect of adding an unhelpable recipient (who is needy). Furthermore, this result rules out added complexity as the reason of increased selfishness in 3PG(0). Oberholzer-Gee and Eichenberger (2008) find that adding other options alongside a dictator game—increasing its complexity—increases selfishness. Furthermore, Rand et al., (2012) find that increasing deliberation results in slower decisions and more selfishness. However, given 3PG(10) has the same level of complexity as 3PG(0), where an extra recipient has been added, if complexity was increasing selfish behaviour it should be observed at a similar rate in 3PG(10). Given we find it is significantly lower than 3PG(0) we can rule out complexity as the reason for increased selfishness in 3PG(0).

The results from **H1** and **H2** suggest violations of the axiom of independence of irrelevant alternatives. The only difference between Control and 3PG(0) is the addition of an unhelpable recipient who receives a payoff of 0 no matter what. Similarly, the only difference between 3PG(0) and 3PG(10) is the quantum of the payoff of the unhelpable recipient. We should expect these changes not to affect the decision of the dictator. However, we find that in both cases this affects behavior, providing evidence that participants violate the axiom of independence of irrelevant alternatives in this environment.

Next, we test **H3**. Table 3 below reports behavior in our 3PG(ambiguous) treatment. Recall that participants are either playing 3PG(10) or 3PG(0), but do not know which version they are in. The only difference between these two environments is how much the unhelpable recipient will receive. We find that a minority of participants (22.7%) chose to reveal which version of the 3PG they were in.

Table 3. Summary of choices made in 3PG(ambiguous).

Choices	**% Chosen (n)**
Did not reveal	77.2% (550)
Option 1 (generous)	29.3% (208)
Option 2 (selfish)	47.9% (342)
Reveal	22.7% (161)
Revealed 3PG(10)	
Option 1 (generous)	8.6% (61)
Option 2 (selfish)	2.8% (20)
Revealed 3PG(0)	
Option 1 (generous)	8.2% (58)
Option 2 (selfish)	3.1% (22)

We start by investigating H3a, the behavior in 3PG(ambiguous) of non-discriminating types who choose the same option in 3PG(0) and 3PG(10). Table 4 reports the choices of these non-discriminating types, aggregating together participants who revealed and did not – since for these types revealing should not change behavior. Of the 160 participants who consistently chose Option 2 in both the full information games 3PG(0) and 3PG(10), we find that 90% of them choose Option 2 in 3PG(ambiguous). Of the 409 participants who consistently chose Option 1, we find that 34.7% of them change their choice in 3PG(ambiguous) and choose option 2. The rate of selfish behavior significantly increases in 3PG(ambiguous) compared to 3PG(0) and 3PG(10) for these types (50.3% vs 28.2%, McNemar test, $p<0.001$). These results suggest that a significant number of participants who choose Option 1 in both games, select the selfish option in the presence of uncertainty. We thus find evidence for **H3A,** that is a difference in the rate of selfish behavior in Games 3PG(0) and 3PG(10) compared to 3PG(ambiguous) for those who choose the same option in the full information treatments.

Table 4. Choices made in 3PG(ambiguous) 3PG Moral Wiggle Room Game for Non-Discriminating Types in 3PG(0) and 3PG(10).

	Option 1 (3PG(ambiguous))	Option 2 (3PG(ambiguous))	Total
O1 in 3PG(0)/3PG(10)	267	142	409
O2 in 3PG(0)/3PG(10)	16	144	160
Total	283	286	569

Next, we test **H3B**. This hypothesis applies to participants who behave inconsistently in 3PG(0) and 3PG(10), and who reveal in 3PG(ambiguous). Our null hypothesis is that participants will behave consistently with their behavior in whichever game they have revealed themselves to be in. We have very few observations in this category. Although 10 out of 21 cases do choose inconsistently with their choice in the simple treatments, the inconsistencies are distributed perfectly evenly across the two directions, thus we cannot reject the null that they are choosing selfish or generous behavior with the same probability across the two games ($p = 1.00$, McNemar test).

Finally, for **H3c** we test whether the rate of selfishness for discriminating types that do not reveal in 3PG(ambiguous) differs from 50%. Recall that because they chose Option 1 in 3PG(0) or 3PG(10), and Option 2 in the other, if they are simply randomizing their choice, the rate of selfishness should not be different to 50%. We find that the rate of selfishness in 3PG(ambiguous) for discriminatory types who do not reveal is 73.6%. This is statistically significant from the random choice of 50% (binomial test, $p<0.0001$).

We conduct further analysis on **H3c** by looking at the discriminating types who conform to compassion fade – meaning that they choose selfishly in 3PG(0), and generously in 3PG(10). We find that 77% of these discriminators who don't reveal in 3PG(ambiguous) choose the selfish option. A binomial test confirms that this is significantly different from a mixed strategy of choosing each option with 50% probability ($p<0.0001$).

These results suggest that people use the lack of information to justify acting in their own self-interest. This supports the guilt mitigation account of compassion fade. In the full information games, participants appear to suffer greater guilt from choosing selfishly in 3PG(0) than 3PG(10).

4.4.1 Classification of Types

Since participants completed the 5 games (A-E) in the first block, we are able to classify them into different preference structures, defined by these five decisions. The types we use for categorization are explained below.[46]

1. **Competitive:** These types have a lexicographic preference structure where they first look to maximize their own payoff, and then maximize the difference between their payoff and the payoff of the other recipients.
   Choices: This type will choose Option 2 in all games

2. **Narrow Self Interest:** This type maximizes their own payoff and is indifferent to the payoff of others.
   Choices: This type will choose Option 2 in the Control, 3PG(0) and 3PG(10). They are indifferent in NoProfit-Control and NoProfit-3PG(0).[47]

3. **Welfare Maximiser:** This type attempts to maximize the total payoffs of all players combined.
   Choices: This type will choose Option 1 in all games

4. **Weak Compassion Fade:** This type has a preference to maximize welfare, but exhibits compassion fade in the presence of a needy individual who cannot be helped. This results in them choosing selfishly in this situation. Note: Compassion fade only occurs for this type when the unhelpable player is needy.
   Choices: This type chooses Option 1 in Control, 3PG(10), NoProfit-Control and NoProfit-3PG(0). This type chooses Option 2 in 3PG(0)

5. **Strong Compassion Fade:** This type has a preference to maximize welfare, but this motivation diminishes in the presence of others who cannot be helped, regardless of how well off they are.
   Choices: This types chooses Option 1 in Control, NoProfit-Control and NoProfit-3PG(0). This type chooses Option 2 in 3PG(0) and 3PG(10).

[46] The seven categories we propose were all pre-registered before running the experiment.
[47] Note: If they choose Option 2 in both games D and E we count them as competitive types, even though they could technically be narrow self-interested types. This is to avoid double counting.

6. **Inequality Averse 1**: This type has a lexicographic preference to minimize the inequality between recipients, then looks to maximize the total payoffs of all players combined.

   Choices: This type chooses Option 2 in 3PG(0) and NoProfit-3PG(0). This type chooses Option 1 in Control, 3PG(10) and NoProfit-Control.

7. **Inequality Averse 2:** This type has a lexicographic preference to minimize the inequality between recipients, then looks to maximize their own payoff

   Choices: This type chooses Option 2 in Control, 3PG(0), NoProfit-3PG(0). This type chooses Option 1 in 3PG(10). This type is indifferent in NoProfit-Control.

We can assign 77.4% of our participants into these 7 categories. We present the breakdown in Table 5 below. The majority of classified participants are welfare maximisers, with 47.3% of the total sample having these preferences. We find that 5.3% of participants fall into the Weak Compassion Fade category and 3.1% fall into the Strong Compassion Fade category. The behavior of both these types are affected by the addition of an unhelpable player, with the former's behavior differing with the level of the unhelpable players' neediness. This shows that there are some participants (8.4%) who are affected by the addition of a dummy player in dictator game environments. We also find that very few participants have preference structures that focus on inequality aversion between recipients, with only 3.2% of participants falling into the categories of Inequality Averse 1 and Inequality Averse 2. These results support our previous analysis that the behavioral anomalies that occur in 3PG(0) and 3PG(10) can be more readily explained by a compassion fade mechanism, rather than as a result of distributional preferences.

Table 5. Classification and Frequency of Types.

Type	**Frequency (%)**
Competitive	1.4
Narrow Self-Interest	17.0
Welfare Maximiser	47.3
Weak Compassion Fade	5.3
Strong Compassion Fade	3.1

Inequality Averse 1	1.5
Inequality Averse 2	1.7
Not Classified	22.6

Revisiting our analysis of 3PG(ambiguous), we hypothesize that if compassion fade mitigates feeling of guilt, then we should observe differences between Weak/Strong Compassion Fade types on the one hand, and Welfare Maximisers on the other. In Table 6 below, we report the behavior in 3PG(ambiguous), conditional on preference type. We find – unsurprisingly – that the majority of participants classified with selfish preferences (Weak Self Interest and Competitive types) choose selfishly. Even when they reveal the state of the world, 95% (21 of 22) still choose the selfish option. Conversely, we find a plurality of welfare maximisers (38%) do not reveal and choose Option 1, with 30.7% not revealing and choosing Option 2. When these types reveal what state they are in (31.0%) they choose the generous option 90.4% of the time. In contrast to welfare maximisers, we find that our Weak/Strong Compassion faded types do not reveal and choose Option 2 51.7% of the time. This is significantly more frequently than Welfare Maximizers ($X^2 = 9.1$, $p = 0.003$). This shows that types with Compassion Faded preferences act differently (more selfishly) than Welfare Maximisers in an ambiguous information environment. This provides further evidence for the mechanism of compassion fade being a guilt mitigation account rather than a decrease in altruistic effect. If it were the latter, we would expect to see similar behavior between welfare maximisers and Compassion Faded types.

Table 6. Decision in 3PG(ambiguous) conditional on type.

Choices	**Weak/Strong Compassion Fade (n)**	**Welfare Maximiser (n)**	**Weak Self Interest/Competitive (n)**
Did not reveal	75.0% (n=45)	69.1% (n=232)	83.2% (n=109)
Option 1 (generous)	23.3% (14)	38.4% (129)	4.6% (6)
Option 2 (selfish)	51.7% (31)	30.7% (103)	78.6% (103)
Reveal	25.0% (15)	29.9% (104)	16.8% (22)

5. Discussion

Our second experiment has confirmed the existence of compassion fade. However, the magnitude of the effect is greatly reduced. This may be a result of the within-subject design where participants can detect inconsistencies in their choices. We found that approximately 8.4% of participants manifested some variety of compassion fade – reducing their generosity either when the number of unhelpable individuals in the decision frame increased, or when the wellbeing of the unhelpable individuals was lowered. We also confirmed that both elements of compassion fade were relevant: the number of unhelpable individuals and the neediness of the unhelpable individuals both affected the rate of generosity. We also observed that these results could not be accounted for in terms of an aversion to creating inequality between the helpable and unhelpable recipients.

Regarding the increase in compassion fade when the unhelpable recipient was worse off, our evidence does not allow us to make any strong claims about the mechanism involved here. However, we suggest a plausible intuition is that when a decision maker is reminded that helping one person will still leave others suffering, it is easier to mitigate feelings of guilt for not helping – “the problem is too big for me to solve”; “my action is just a drop in the bucket”, etc. But if the unhelpable recipient is better off, then it instead reinforces the salience of the deprivation which the decision maker can alleviate.[48] Future studies could try and identify whether it is the relative, or absolute level of neediness that is driving this effect.

The results from our treatment with incomplete information support the hypothesis that some individuals are able to self-induce compassion fade to act egoistically. For both our non-discriminating and discriminating types in 3PG(ambiguous), we find a substantial increase in the frequency of selfish behavior from those who refuse the opportunity to be fully informed.

These results provide some evidence that compassion fade operates via the mechanism of reducing the guilt one feels for not acting generously. If compassion fade primarily diminished warm glow, then participants would either choose not to reveal and select the generous option or reveal the true state of the world before making their decisions. However, as our results show, this is not the case. Participants who do not reveal the information are significantly more likely to choose the selfish option. This is suggestive that participants who can manipulate their beliefs are doing so in order to mitigate a negative utility factor, such as guilt, and hence that guilt mitigation is the primary mechanism of compassion fade.

[48] This can also be explained by the reference point that the unhelpable recipient’s payoff provides the decision maker.

Regarding our interpretation in terms of guilt mitigation, our results bear on a number of earlier studies. Contrary to our results, Erlandsson et al. (2016) interpret their observations of compassion fade in terms of warm glow diminution. They ask their participants whether they expect to feel warm glow if they give. In two of their experiments, they ask participants to report on their emotions regarding a possible donation decision. One experiment explicitly asks for forecasts of anticipated warm glow, but it is not clear that this is relevant to predicting actual behavior. The other experiment asks to what extent participants were “touched” by the donation request, but this obscures the difference between positive affect for generous behavior and negative affect for selfish behavior. In the only experiment with a clear measure of actual giving behavior, they find no effect.

Cameron and co-authors, (2016; 2011) find evidence in favor of the mitigation of guilt. In Cameron and Payne (2011), they find that participants who anticipate being asked to donate to others experience a drop off in empathy that is greater for large numbers of recipients rather than small numbers. In (Cameron et al., 2016), they show a particular pathway via which this drop in empathy might occur: by dehumanization, and they interpret this as evidence that compassion fade works to reduce negative affect for selfish behavior.

Västfjäll et al. (2014) claim they find evidence that is not entirely consistent with the Cameron et al. approach. They test for correlations between the number of victims in the decision problem and participants’ levels of affect. They find that there is less positive affect in scenarios where people are confronted with many victims rather than few. This indeed suggests that warm glow is diminished in the presence of more potential recipients. But even if affect is predictive of behavior, we need to know counterfactually, what would the individual affect be if they made the alternative decision?

An alternative explanation for our results in the 3PG(ambiguous) treatment is that participants may be engaging in moral licensing (Blanken et al., 2015; Mullen & Monin, 2016). This is a phenomenon whereby individuals who have made a virtuous decision earlier will later feel entitled to engage in less virtuous behavior. Participants only encountered 3PG(ambiguous) in block 2, after completing at least 6 tasks in block 1. Having already behaved more virtuously in block 1, perhaps some individuals felt they could ignore the information about the wellbeing of the unhelpable individual, and then choose selfishly. This is an alternative explanation as to why non-discriminating types (choose Option 1 in games B and C) change their decision to Option 2 in 3PG(ambiguous). However, this does not explain why they choose not to reveal the true state of the world. If the moral licensing hypothesis was correct, we should expect to find a high rate of selfish behavior, even if they chose to reveal the true state. However, our data shows that less than 25% of non-discriminating types who revealed chose option 2. Furthermore, we would expect moral licensing not just to affect compassion faded types, but also welfare maximisers. However, we find significant differences in their behavior in 3PG(ambiguous) which suggests that moral licensing is not the main cause of this result. In fact, given that welfare maximisers have done even more good than compassion faded types up until 3PG(ambiguous) we may even expect more selfish behavior from welfare maximizers if moral licensing were driving this result.

6. Conclusion

In this paper we reported the results of two experiments investigating the role that compassion fade plays in altruistic decision making. Previous work has focused on compassion fade in charitable donation settings, where recipients are in dire need and the opportunity to assist is mediated by a charitable organization. We document a similar phenomenon in decisions to allocate money to relative peers in typical economic laboratory settings. When an “unhelpable recipient” is added to the decision frame, even though this makes no difference to the material payoffs that can be affected by the decision maker, there is a significant drop in altruistic behavior. We further demonstrate that this effect depends on the payoff of the unhelpable recipient. When that unhelpable individual receives a relatively generous allocation, we observed a smaller drop in altruistic behavior.

As well as these basic findings regarding the existence of compassion fade, our experiments provide some suggestive evidence regarding its mechanism. If we are correct that participants use an opportunity to be ignorant to manipulate their credences in a utility-favoring fashion, then our findings suggest that participants are motivated to induce compassion fade because it reduces the negative utility associated with selfish behavior.

There are some limitations to our experimental design which suggest some avenues for future research. For example, we only considered decision frames in which there were zero or one unhelpable individuals. It would be interesting to obtain quantitative evidence for the rate at which compassion fade is induced as the number of unhelpable individuals increases further.

References

Almaatouq, A., Krafft, P., Dunham, Y., Rand, D. G., & Pentland, A. (2020). Turkers of the world unite: Multilevel in-group bias among crowdworkers on amazon mechanical turk. *Social Psychological and Personality Science*, *11*(2), 151–159.

Bartels, D. M. (2006). Proportion dominance: The generality and variability of favoring relative savings over absolute savings. *Organizational Behavior and Human Decision Processes*, *100*(1), 76–95. https://doi.org/10.1016/j.obhdp.2005.10.004

Battigalli, P., & Dufwenberg, M. (2007). Guilt in Games. *American Economic Review*, *97*(2), 170–176. https://doi.org/10.1257/aer.97.2.170

Benabou, R., & Tirole, J. (2002). Self-Confidence and Personal Motivation. *The Quarterly Journal of Economics*, *117*(3), 871–915. https://doi.org/10.1162/003355302760193913

Blanken, I., van de Ven, N., & Zeelenberg, M. (2015). A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin*, *41*(4), 540–558.

Bolsen, T., & Shapiro, M. A. (2018). The US News Media, Polarization on Climate Change, and Pathways to Effective Communication. *Environmental Communication*, *12*(2), 149–163. https://doi.org/10.1080/17524032.2017.1397039

Bolton, G. E., Katok, E., & Zwick, R. (1998). Dictator game giving: Rules of fairness versus acts of kindness. *International Journal of Game Theory*, *27*(2), 269–299. https://doi.org/10.1007/s001820050072

Bolton, G. E., & Ockenfels, A. (2008). Chapter 59 Self-centered Fairness in Games with More Than Two Players. In *Handbook of Experimental Economics Results* (Vol. 1, pp. 531–540). Elsevier. https://doi.org/10.1016/S1574-0722(07)00059-5

Butts, M. M., Lunt, D. C., Freling, T. L., & Gabriel, A. S. (2019). Helping one or helping many? A theoretical integration and meta-analytic review of the compassion fade literature. *Organizational Behavior and Human Decision Processes*, *151*, 16–33. https://doi.org/10.1016/j.obhdp.2018.12.006

Cameron, C. D., Harris, L. T., & Payne, B. K. (2016). The Emotional Cost of Humanity: Anticipated Exhaustion Motivates Dehumanization of Stigmatized Targets. *Social Psychological and Personality Science*, *7*(2), 105–112. https://doi.org/10.1177/1948550615604453

Cameron, C. D., & Payne, B. K. (2011). Escaping affect: How motivated emotion regulation creates insensitivity to mass suffering. *Journal of Personality and Social Psychology*, *100*(1), 1–15. https://doi.org/10.1037/a0021643

Capra, C. M., & Larson, T. (2009). Exploiting moral wiggle room: Illusory preference for fairness? A comment. *Judgment and Decision Making*, *4*, 467–474.

Capraro, V., & Perc, M. (2021). Mathematical foundations of moral preferences. *Journal of The Royal Society Interface*, *18*(175), 20200880. https://doi.org/10.1098/rsif.2020.0880

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97. https://doi.org/10.1016/j.jbef.2015.12.001

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*(1), 67–80. https://doi.org/10.1007/s00199-006-0153-z

Di Tella, R., Perez-Truglia, R., Babino, A., & Sigman, M. (2015). Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others' Altruism. *American Economic Review*, *105*(11), 3416–3442. https://doi.org/10.1257/aer.20141409

Dinno, A. (2017). Tost: Two one-sided tests for equivalence. *Stata Software Package*.

Drews, S., & van den Bergh, J. C. J. M. (2016). What explains public support for climate policies? A review of empirical and experimental studies. *Climate Policy*, *16*(7), 855–876. https://doi.org/10.1080/14693062.2015.1058240

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, 643–669.

Erlandsson, A., Björklund, F., & Bäckström, M. (2014). Perceived Utility (not Sympathy) Mediates the Proportion Dominance Effect in Helping Decisions: Perceived Utility Mediates the PDE. *Journal of Behavioral Decision Making*, *27*(1), 37–47. https://doi.org/10.1002/bdm.1789

Erlandsson, A., Västfjäll, D., Sundfelt, O., & Slovic, P. (2016). Argument-inconsistency in charity appeals: Statistical information about the scope of the problem decrease helping toward a single identified victim but not helping toward many non-identified victims in a refugee crisis context. *Journal of Economic Psychology*, *56*, 126–140. https://doi.org/10.1016/j.joep.2016.06.007

Exley, C., & Kessler, J. (2019). *Motivated Errors* (No. w26595; p. w26595). National Bureau of Economic Research. https://doi.org/10.3386/w26595

Exley, C. L. (2016a). Excusing Selfishness in Charitable Giving: The Role of Risk. *The Review of Economic Studies*, *83*(2), 587–628. https://doi.org/10.1093/restud/rdv051

Exley, C. L. (2016b). *Using charity performance metrics as an excuse not to give*.

Exley, C. L. (2019). Using charity performance metrics as an excuse not to give. *Management Science*, mnsc.2018.3268. https://doi.org/10.1287/mnsc.2018.3268

Fisman, R., Kariv, S., & Markovits, D. (2007). Individual Preferences for Giving. *American Economic Review*, *97*(5), 1858–1876. https://doi.org/10.1257/aer.97.5.1858

Gandullia, L. (2019). The price elasticity of warm-glow giving. *Economics Letters*, *182*, 30–32.

Gandullia, L., & Lezzi, E. (2018). The price elasticity of charitable giving: New experimental evidence. *Economics Letters*, *173*, 88–91.

Gandullia, L., Lezzi, E., & Parciasepe, P. (2020). Replication with MTurk of the experimental design by Gangadharan, Grossman, Jones & Leister (2018): Charitable giving across donor types. *Journal of Economic Psychology*, *78*, 102268.

Gino, F., Norton, M. I., & Weber, R. A. (2016). Motivated Bayesians: Feeling moral while acting egoistically. *The Journal of Economic Perspectives*, *30*(3), 189–212. https://doi.org/10.1257/jep.30.3.189

Glynn, A. N. (2013). What Can We Learn with Statistical Truth Serum?: Design and Analysis of the List Experiment. *Public Opinion Quarterly*, *77*(S1), 159–172. https://doi.org/10.1093/poq/nfs070

Grossman, Z., & der Weele, J. J. V. (2013). Self-image and strategic ignorance in moral dilemmas. *Available at SSRN 2237496*. http://papers.ssrn.com/sol3/papers.cfm?abstract{_}id=2237496

Hamman, J. R., Loewenstein, G., & Weber, R. A. (2010). Self-Interest through Delegation: An Additional Rationale for the Principal-Agent Relationship. *American Economic Review*, *100*(4), 1826–1846. https://doi.org/10.1257/aer.100.4.1826

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400–407.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 65–70.

Horne, Z., & Livengood, J. (2017). Ordering effects, updating effects, and the specter of global skepticism. *Synthese*, *194*(4), 1189–1218. https://doi.org/10.1007/s11229-015-0985-9

Kahneman, D., & Tversky, A. (1996). *On the reality of cognitive illusions.*

Kassas, B., & Palma, M. A. (2019). Self-serving biases in social norm compliance. *Journal of Economic Behavior & Organization*, *159*, 388–408. https://doi.org/10.1016/j.jebo.2019.02.010

Kopec, M., & Bruner, J. (2021). No Harm Done? An Experimental Approach to the Nonidentity Problem. *Journal of the American Philosophical Association*.

Kripke, S. A. (1972). Naming and necessity. In *Semantics of natural language* (pp. 253–355). Springer.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Loewenstein, G., & Thaler, R. H. (1989). Anomalies: Intertemporal Choice. *Journal of Economic Perspectives*, *3*(4), 181–193. https://doi.org/10.1257/jep.3.4.181

Loomes, G. (1991). Evidence of a New Violation of the Independence Axiom. *Journal of Risk and Uncertainty*, *4*(1), 91–108.

Machina, M. J. (1987). Choice Under Uncertainty: Problems Solved and Unsolved. *Journal of Economic Perspectives*, *1*(1), 121–154. https://doi.org/10.1257/jep.1.1.121

Markowitz, E., Slovic, P., Vastfjall, D., & Hodges, S. (2013). *Compassion fade and the challenge of environmental conservation*. https://scholarsbank.uoregon.edu/xmlui/handle/1794/22102

Masson, T., & Fritsche, I. (2021). We need climate change mitigation and climate change mitigation needs the 'We': A state-of-the-art review of social identity effects motivating climate change action. *Current Opinion in Behavioral Sciences*, *42*, 89–96. https://doi.org/10.1016/j.cobeha.2021.04.006

Matthey, A., & Regner, T. (2011). Do I Really Want to Know? A Cognitive Dissonance-Based Explanation of Other-Regarding Behavior. *Games*, *2*(1), 114–135. https://doi.org/10.3390/g2010114

Mullen, E., & Monin, B. (2016). Consistency versus licensing effects of past moral behavior. *Annual Review of Psychology*, *67*.

Nielsen, K., & Rehbeck, J. (2022). When choices are mistakes. *American Economic Review*, *112*(7), 2237–2268.

Oberholzer-Gee, F., & Eichenberger, R. (2008). Fairness in extended dictator game experiments. *The BE Journal of Economic Analysis & Policy*, *8*(1).

Palan, S., & Schitter, C. (2018). Prolific. Ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27.

Parfit, D. (1984). *Reasons and Persons*. OUP Oxford.

Raihani, N. J., Mace, R., & Lamba, S. (2013). The Effect of $1, $5 and $10 Stakes in an Online Dictator Game. *PLOS ONE*, *8*(8), e73131. https://doi.org/10.1371/journal.pone.0073131

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, *489*(7416), Article 7416. https://doi.org/10.1038/nature11467

Saccardo, S., & Serra-Garcia, M. (2020). Cognitive Flexibility or Moral Commitment? Evidence of Anticipated Belief Distortion. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3676711

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, *15*(6), 657–680.

Serra-Garcia, M., & Szech, N. (2019). *The (in) elasticity of moral ignorance*.

Shalvi, S., Dana, J., Handgraaf, M. J. J., & De Dreu, C. K. W. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes*, *115*(2), 181–190. https://doi.org/10.1016/j.obhdp.2011.02.001

Slovic, P. (2007). "If I look at the mass I will never act": Psychic numbing and genocide. *Judgment and Decision Making*, *2*(2), 79–95.

Small, D. A., Loewenstein, G., & Slovic, P. (2007). Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behavior and Human Decision Processes*, *102*(2), 143–153. https://doi.org/10.1016/j.obhdp.2006.01.005

Sznycer, D., Xygalatas, D., Agey, E., Alami, S., An, X.-F., Ananyeva, K. I., Atkinson, Q. D., Broitman, B. R., Conte, T. J., Flores, C., Fukushima, S., Hitokoto, H., Kharitonov, A. N., Onyishi, C. N., Onyishi, I. E., Romero, P. P., Schrock, J. M., Snodgrass, J. J., Sugiyama, L. S., … Tooby, J. (2018). Cross-cultural invariances in the architecture of shame. *Proceedings of the National Academy of Sciences*, 201805016. https://doi.org/10.1073/pnas.1805016115

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Västfjäll, D., Slovic, P., Mayorga, M., & Peters, E. (2014). Compassion Fade: Affect and Charity Are Greatest for a Single Child in Need. *PLOS ONE*, *9*(6), e100115. https://doi.org/10.1371/journal.pone.0100115

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*(1), 75–98. https://doi.org/10.1007/s10683-009-9230-z

Appendix

Appendix A: Experiment 1: Original Motivation and Full Design

Experiment 1 was originally conceived as a way to examine one possible barrier to generating altruistic behavior, when the desired altruistic acts benefit people in the distant future. Take climate change, for example: it has proven extremely difficult to motivate the general public to support policies that mitigate against climate change – even policies that involve relatively minor sacrifices – when the effects of *not* making those sacrifices could be absolutely disastrous for future people. A number of explanations have been offered for this reluctance in the literature (Drews & van den Bergh, 2016; Bolsen & Shapiro, 2018; Masson & Fritsche, 2021), but one underexplored possibility stems from a philosophical puzzle known as the nonidentity problem (Parfit, 1984; Kopec & Bruner, 2021). This puzzle stems from the fact that some of our actions or policies not only affect the welfare of future people, they also affect exactly which people end up coming into existence.

Effective climate policy will arguably need to have substantial effects on people's lives, by, for example, clustering populations closer together, making work commutes shorter (or non-existent), raising the cost of long-distance travel, changing what people do for recreation, modifying what people eat, encouraging people to have fewer children, etc. Although it may not be immediately obvious, the choice to enact a policy like this will, in due time, result in a future population that contains a completely different group of people than would have existed had the policy not been enacted. Some people will meet (and eventually procreate with) people they wouldn't have otherwise met, e.g., meeting on public transportation when they would have otherwise driven. Some people who would have existed under business-as-usual now won't exist, as the population is encouraged to contract. Other people will wait longer to have children, as roomier suburban housing becomes more expensive. Some philosophers have argued that the same person could not have been created from a different pair of sperm and egg cells (Kripke, 1972). If that is correct, then a procreative delay of roughly a month is sufficient to guarantee that a different human will exist, and even a much smaller disruption will very likely have the same effect. And even on more relaxed views about the nature of personal identity, it is overwhelmingly plausible that the aggregate effects of the substantial measures required to address climate change, over a long enough timeframe, would cause a completely different population to exist than would have existed had the policy not been enacted.

Moral philosophers have their own reasons for finding the above metaphysical quirk theoretically puzzling (see Kopec & Bruner, 2021) , but our concern was more practical. We worried that if the public could grasp the identity-affecting nature of policies like this, that could exacerbate the public's hesitance toward supporting policies that require some sacrifices. In other words, people may become less altruistic when facing identity-affecting choices. After all, those future people would have a hard time claiming that *they* have been harmed by our choice not to enact the policy, because if we *had* enacted the policy, *they* would not have existed. Some other group of people would be enjoying a planet with lower greenhouse gas concentrations, not *them*.

But this would be a very difficult question to probe directly, since one would need to devise an experiment that actually changes which people come into existence. Although such experiments are conceivable, they aren't feasible. So we devised a proxy for non-existence. Arguably, part of what moves people to make altruistic sacrifices, e.g., in dictator games, is that the other party affected by the decision would know if a selfish choice was made. This suggests that we might be able to use ignorance of the decision as a proxy for a lack of existence. This then led us to the experimental design. We created a version of the classic dictator game, but with two possible recipients, one of which is guaranteed not to know anything at all about the experiment. The choice is binary, between an altruistic split and a selfish split, and the choice between being altruistic or selfish not only affects the pay-off of the dictator and the recipient, it also affects which recipient gets the reward and, in turn, which "recipient" ends up completely ignorant of the experiment as a whole. This set-up was intended to mimic the motivational quirk mentioned above. Suppose a dictator makes the selfish choice. The resulting recipient would have a hard time claiming that they have been harmed by the dictator's selfish choice. Afterall, if the dictator had chosen the altruistic choice, *this recipient* wouldn't have received that higher bonus - it would have gone to the other recipient. *This recipient* wouldn't even know about the experiment. As far as the experimental setup is concerned, they might as well have not existed if the other choice had been made.

All the treatments we used are described below in Table A1. To summarize, we found that all three interventions resulted in significantly more selfish behavior than the Control, but there were no differences between 3PG(0) and either the Identity-Affecting Treatment, or Forced-Tradeoff treatment. This result suggests that the increase in selfish behavior is mostly due to the addition of a dummy recipient who receives $0 in 3PG($0). This is what motivated us to conduct Experiment 2.

Table A1: All Treatments in Experiment 1

Treatment	Options
Control	A: (10,10) B: (12,2)
Identity-Affecting Treatment	A: (10,10, Ω) B: (12, Ω,2)
Forced-Tradeoff Treatment	A: (10,10,0) B: (12,0,2)
3PG(0)	A: (10,10,0) B: (12,2,0)

Notes: In the Identity-Affecting Treatment, Ω means the recipient receives $0, but never finds out they were a part of the experiment.

Appendix B: Experiment 1 Instructions:

1. Task 1:

[Control Treatment]

In this task, you have been randomly matched with a unique individual (no one else has been matched with this person). You and this other participant will remain mutually anonymous. The participant you have been randomly matched with is not present at the moment. We will invite them to a venue on campus next week. We will explain to these participants the procedure for the experiment you are doing now, so they understand the choices you made, and how their payment was determined.

If you wish to check on our procedure, you may come to the venue next week to observe the process and confirm that we make these payments as promised. It is very important to us that you realise that the decision you make has real consequences for yourself and for another participant.

You need to make a decision about how to allocate some money between yourself and the other participant. The money has been initially allocated so that both of you receive $10. You need to choose between two options (pictured below).

Option 1: You keep the current allocations: you receive $10, the other participant receives $10.

Option 2: You can elect to take an extra $2 for yourself, in which case the other participant's allocation will be reduced to $2. So, you will receive $12, and they will receive $2.

You Choose	You Receive	Other Participant Receives
Option 1	$10	$10
Option 2	$12	$2

[3PG(0) Treatment]

In this task, you have been randomly matched with two unique individuals (no one else has been matched with either of these people) Participant A and Participant B. You, Participant A and Participant B will remain mutually anonymous.

Participant A and Participant B are not present at the moment. Depending on your decision, we will invite one of them to a venue on campus next week. We will explain to the invited participants the procedure for the experiment you are doing now, so they understand the choices you made, and how their payment was determined.

It is very important to us that you realise that the decision you make has real consequences for yourself, Participant A and Participant B.

You need to make a decision about how to allocate some money between yourself and the other participants. The money has been initially allocated so that you receive $10, Participant A receives $10 and Participant B receives $0. You need to choose between two options (pictured below).

Option 1: You keep the current allocations: you receive $10, Participant A receives $10 and Participant B receives $0

Option 2: You can elect to take an extra $2 for yourself, in which case Participant A's allocation will be reduced to $2. So, you will receive $12, Participant A will receive $2 and Participant B will receive $0.

Note: We will only invite a participant to the venue next week if they are to receive some money as a result of your decision. If a participant will receive $0 as a result of your decision, we will not invite them to the venue, and they will never learn about the experiment.

You Choose	You Receive	Participant A Receives	Participant B Receives
Option 1	$10	$10	$0
Option 2	$12	$2	$0

***At this point, to make sure that everyone understands the game, please answer the two questions below*

1. If You choose Option 2 then You receive $___, Participant A receives $___, Participant B receives $__ ($12, $2, $0)

2. If You Choose Option 1 then You receive $___, Participant A receives $___, Participant B receives $__ ($10, $10, $0)

3. If you choose Option A or B, then will Participant B *ever be informed about this experiment? Y/N. (Y)*

Task 2:

(In the following task subjects will be randomised into two groups within the session).

(Group 1)

How many of the following four statements do you agree with?

This laboratory has been operating for over twenty years

This research project is funded by the ARC

The lead researcher of this project obtained his PhD in the USA

Today's experiment was intended to study the psychological basis of behavior that is relevant to environmental conservation policy

(Group 2)

How many of the following four statements do you agree with?

This laboratory has been operating for over twenty years

This research project is funded by the ARC

The lead researcher of this project obtained his PhD in the USA

Today's experiment was intended to study the psychological basis of behaviour that is relevant to environmental conservation policy

The recipients in Task One will be paid in the way described in the instructions.

Task 3:

(Subjects are given 9 questions to answer on a 7-point Likert scale. For each option they can answer, Strongly Disagree, Disagree, Somewhat Disagree, Neither Agree Nor Disagree, Somewhat Agree, Agree, Strongly Agree.)

Please read the following list of statements and indicate how much you agree or disagree with each statement:

1. If the only way to save another person's life during an emergency is to sacrifice one's own leg, then one is morally required to make this sacrifice.
2. From a moral point of view, we should feel obliged to give one of our kidneys to a person with kidney failure since we don't need two kidneys to survive, but really only one to be healthy.
3. From a moral perspective, people should care about the well-being of all human beings on the planet equally; they should not favour the well-being of people who are especially close to them either physically or emotionally.
4. It is just as wrong to fail to help someone as it is to actively harm them yourself.
5. It is morally wrong to keep money that one doesn't really need if one can donate it to causes that provide effective help to those who will benefit a great deal.
6. It is morally right to harm an innocent person if harming them is a necessary means to helping several other innocent people.
7. If the only way to ensure the overall well-being and happiness of the people is through the use of political oppression for a short, limited period, then political oppression should be used.
8. It is permissible to torture an innocent person if this would be necessary to provide information to prevent a bomb going off that would kill hundreds of people.
9. Sometimes it is morally necessary for innocent people to die as collateral damage—if more people are saved overall.

Appendix C: Experiment 2 All Treatments

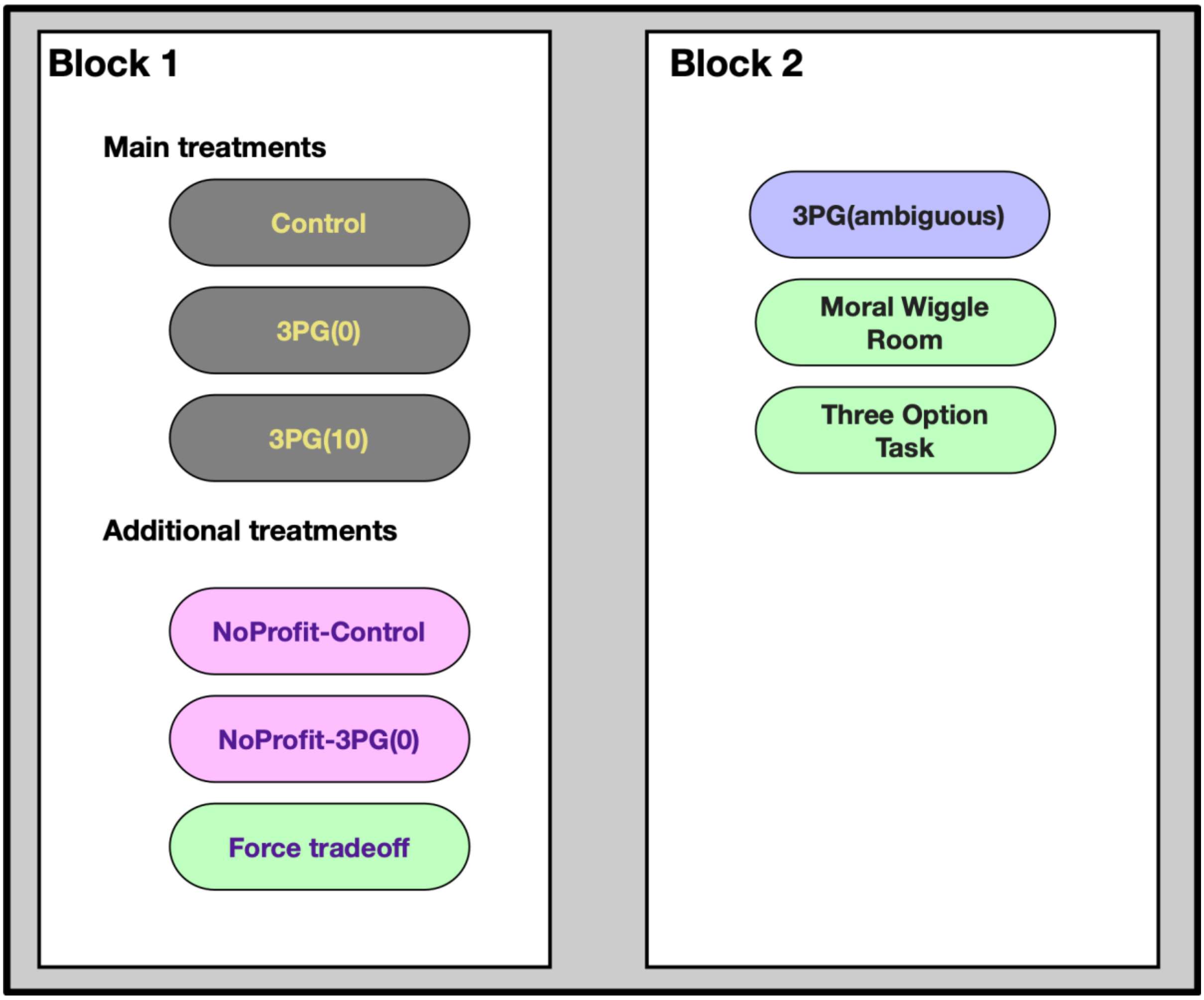


Figure C1: The three tasks in green were omitted from Figure 5 in the main paper. The Forced Tradeoff task is the same as in Experiment 1 (See Appendix A). The Moral Wiggle Room Task is the same as in (Dana et al., 2007). Finally, in the Three Option task, the dictator could pick Option 1 (10, 10, 0), Option 2 (10, 0, 2) or Option 3 (13, 3, 0). This was all captured in our pre-registration.

Appendix D: Experiment 2 Instructions

You are about to perform a number tasks involving one or more other participants in this overall study. Each task is independent of the other, meaning your decision in the first task has no effect on the second task, and vice versa. At the end of the experiment only one of the tasks will be chosen for payment. At the end of the experiment, we will use an automated, random procedure to choose only one of the tasks for payment. Both tasks are equally likely to be chosen.

In each of the tasks to follow, you will be making a choice that will affect either one or two other participants. All participants will remain mutually anonymous.

In preparation for this experiment, we previously recruited a bank of MTurk workers to serve as possible payment recipients. We have placed these participants into a random sequence (a “queue”). As soon as you successfully complete all aspects of the experiment, you will be matched with the relevant number of people who are next up in this queue (that is, either the next individual in line or the next two individuals in line, depending on which task is chosen). No one else will be matched with that person or those people, so your choice is the only one that will affect their payoff from this study.

After you complete the experiment, and we have chosen which task will be paid, we will send, along with any payment, a note to the recipient(s) explaining the procedure for the relevant task, so they understand the choice you made, and how their payment was determined. The recipients will not be given the opportunity to participate in the study as decision-makers, and you (a decision maker) will not be given the opportunity to later participate as a recipient.

It is very important to us that you realize that the decision you make has real consequences for yourself and one or more other workers. We have also reported our experimental procedure to our Human Research Ethics Board, and they will monitor that we complete the recipient payments as we have specified in this description.

In these tasks, monetary amounts are not quoted in USD, but rather in Experimental Currency Units (ECUs). Eventually, the amount of money earned during the experiment will be converted into USD, where: **1 ECU= $0.2 USD.**

Note that the tasks will be presented in a random order.

Task A:

In this task, some Experimental Currency Units (ECUs) has been initially allocated so that you and one other recipient each receive 10 ECUs. You need to choose between two options (pictured below).

Option 1: You keep the current allocations: you receive 10 ECUs, the other participant receives 10 ECUs.

Option 2: You can elect to take an extra 2 ECUs for yourself, in which case the other participant's allocation will be reduced to 2 ECUs. So, you will receive 12 ECUs, and they will receive 2 ECUs.

You choose:	You receive	Recipient Receives
Option 1	10	10
Option 2	12	2

Task B:

In this task, the Experimental Currency Units (ECUs) have been initially allocated so that you receive 10 ECUs, Recipient A receives 10 ECUs and Recipient B receives 0 ECUs. You need to choose between two options (pictured below).

Option 1: You keep the current allocations: you receive 10 ECUs, Recipient A receives 10 ECUs and Recipient B receives 0 ECUs.

Option 2: You can elect to take an extra 2 ECUs for yourself, in which case Recipient A's allocation will be reduced to 2 ECUs, but Recipient B's allocation stays the same. So, you will receive 12 ECUs, Recipient A will receive 2 ECUs and Recipient B will receive 0 ECUs.

Note: If a possible recipient receives 0 ECUs as a result of your decision, they will still be made aware of their involvement in the study, and what decision you made.

You choose:	You receive	Recipient A Receives	Recipient B Receives
Option 1	10	10	0
Option 2	12	2	0

Task C:

In this task, some Experimental Currency Units (ECUs) has been initially allocated so that you and one other recipient each receive 10 ECUs. You need to choose between two options (pictured below).

Option 1: You keep the current allocations: you receive 10 ECUs, the other participant receives 10 ECUs.

Option 2: You still receive 10 ECUs and the other participant's allocation will be reduced to 2 ECUs. So, you will receive 10 ECUs, and they will receive 2 ECUs.

You choose:	You receive	Recipient Receives
Option 1	10	10
Option 2	10	2

Task D:

In this task, the Experimental Currency Units (ECUs) have been initially allocated so that you receive 10 ECUs, Recipient A receives 10 ECUs and Recipient B receives 0 ECUs. You need to choose between two options (pictured below).

Option 1: You keep the current allocations: you receive 10 ECUs, Recipient A receives 10 ECUs and Recipient B receives 0 ECUs.

Option 2: You can elect to take an extra 2 ECUs for yourself, in which case Recipient A's allocation will be reduced to 0 ECUs, and recipient B will gain 2 ECUs. So, you will receive 12 ECUs, Recipient A will receive 0 ECUs and Recipient B will receive 2 ECUs.

Note: If a possible recipient receives 0 ECUs as a result of your decision, they will still be made aware of their involvement in the study, and what decision you made.

You choose:	You receive	Recipient A Receives	Recipient B Receives
Option 1	10	10	0
Option 2	12	0	2

Task E:

In this task, the Experimental Currency Units (ECUs) have been initially allocated so that you receive 10 ECUs, Recipient A receives 10 ECUs and Recipient B receives 0 ECUs. You need to choose between two options (pictured below).

Option 1: You keep the current allocations: you receive 10 ECUs, Recipient A receives 10 ECUs and Recipient B receives 0 ECUs.

Option 2: Your payoff stays the same, Recipient A's allocation will be reduced to 2 ECUs, and Recipient B's allocation stays the same. So, you will receive 10 ECUs, Recipient A will receive 2 ECUs and Recipient B will receive 0 ECUs.

Note: If a possible recipient receives 0 ECUs as a result of your decision, they will still be made aware of their involvement in the study, and what decision you made.

You choose:	You receive	Recipient A Receives	Recipient B Receives
Option 1	10	10	0
Option 2	10	2	0

Task F:

In this task, the Experimental Currency Units (ECUs) have been initially allocated so that you receive 10 ECUs, Recipient A receives 10 ECUs and Recipient B receives 0 ECUs. You need to choose between two options (pictured below).

Option 1: You keep the current allocations: you receive 10 ECUs, Recipient A receives 10 ECUs and Recipient B receives 10 ECUs.

Option 2: You can elect to take an extra 2 ECUs for yourself, in which case Recipient A's allocation will be reduced to 2 ECUs, but Recipient B's allocation stays the same. So, you will receive 12 ECUs, Recipient A will receive 2 ECUs and Recipient B will receive 10 ECUs.

You choose:	You receive	Recipient A Receives	Recipient B Receives
Option 1	10	10	10
Option 2	12	2	10

Task G:

This task involves an element of chance: while you know which choice is best for you, you do not know exactly how your choices will affect the other participants. See the table below, to see the two possible choice situations you may be in. With a 50% probability you will be assigned to version 1. With a 50% probability you will be assigned to version 2.

Notice that the only difference between the versions is the payment to the other participant.

In both versions, you get your highest payment of 12 ECUs by choosing option 2.

In Version 1, choosing option 2 gives the other participant his or her lowest payment of 2 ECUs.

In Version 2, choosing option 2 gives the other participant his or her highest payment of 10 ECUs.

(Similarly, in both versions, if you choose option 1, you will receive your lower payment of 10 ECUs. But in Version 1, option 1 gives the other participant his or her highest payment, and in Version 2, option 1 gives the other participant his or her lowest payment.)

Which of these versions will you actually be assigned? That was determined randomly by the computer at the beginning of the session, with each version being equally likely.

Initially, you will not know which version you have been assigned. However, you can choose to find out by clicking a REVEAL button, prior to making your choice. Note that it is optional to press REVEAL: you may make your choice without knowing the version you have been assigned.

Whether you press REVEAL or not will be private. The other participant will not know if you have clicked the button.

Version 1

You choose:	You receive	Recipient Receives
Option 1	10	10
Option 2	12	2

Version 2

You choose:	You receive	Recipient Receives
Option 1	10	2
Option 2	12	10

Task H:

This task involves an element of chance: while you know which choice is best for you, you do not know exactly how your choices will affect the other participants.

See the table below, to see the two possible choice situations you may be in. With a 50% probability you will be assigned to version 1. With a 50% probability you will be assigned to version 2.

Notice that the only difference between the versions is the payment to Recipient B.

In both versions, you get your highest payment of 12 ECUs by choosing option 2, and this is the lowest payment for Recipient A.

In both versions, you get your lowest payment of 10 ECUs by choosing option 1, and this is the highest payment for Recipient A.

In version 1, Recipient B will receive 0 ECUs, no matter what choice you make.

In version 2, Recipient B will receive 10 ECUs, no matter what choice you make.

Which of these versions will you actually be assigned? That was determined randomly by the computer at the beginning of the session, with each version being equally likely.

Initially, you will not know which version you have been assigned. However, you can choose to find out by clicking a REVEAL button, prior to making your choice. Note that it is optional to press REVEAL: you may make your choice without knowing the version you have been assigned.

Whether you press REVEAL or not will be private. The other participant will not know if you have clicked the button.

Version 1

You choose:	You receive	Recipient A Receives	Recipient B Receives
Option 1	10	10	0
Option 2	12	2	0

Version 2

You choose:	You receive	Recipient A Receives	Recipient B Receives
Option 1	10	10	10
Option 2	12	2	10

Task I:

In this task, the Experimental Currency Units (ECUs) have been initially allocated so that you receive 10 ECUs, Recipient A receives 10 ECUs and Recipient B receives 0 ECUs. You need to choose between two options (pictured below).

Option 1: You keep the current allocations: you receive 10 ECUs, Recipient A receives 10 ECUs and Recipient B receives 0 ECUs.

Option 2: You can elect to take an extra 2 ECUs for yourself, in which case Recipient A's allocation will be reduced to 0 ECUs, and Recipient B will gain 2 ECUs. So, you will receive 12 ECUs, Recipient A will receive 0 ECUs and Recipient B will receive 2 ECUs.

Option 3: You can elect to take an ectra $3 for yourself, in which case Recipient A's allocation will be reduced to 3 ECUs and Recipient B will receive 0 ECUs. So, you will receive 13 ECUs, Recipient A receives 3 ECUs and Recipient B receives 0 ECUs.

You choose:	You receive	Recipient A Receives	Recipient B Receives
Option 1	10	10	0
Option 2	12	0	2
Option 3	13	3	0

Appendix E: Experiment 2: Multiple Hypothesis Adjustments

Since we conducted multiple tests, this increases the likelihood of finding a false positive result. We use the Holm-Bonferroni correction method (Holm, 1979) to adjust our critical alpha levels. Below in Table E1, we report all our critical p-values, in order of magnitude, and the relevant Holm-Bonferroni alpha level. Note that H3 is a conjunction of the 3 subsidiary hypotheses (H3A, H3B, H3C). Accordingly, we can reject the null if we reject any one of the three subsidiary hypotheses, and we take the lowest p-value as an upper bound on the p-value of the conjunction.

Table E1: Holm-Bonferroni Multiple Hypothesis Testing

Hypothesis	**P value**	**Adjusted alpha**
H3 (3PG(ambiguous))	$p<0.000001$	0.01
H4 (Forced_Tradeoff)	$p<0.0001$	0.0125
H2 (3PG(0)vs3PG(10))	$p=0.0008$	0.0167
H1(Control vs 3PG(0))	$p=0.028$	0.025
H5 (Third)	$p=1$	0.05

Thus, for having controlled the critical alpha level to be 0.05, we are able to reject the null hypothesis for H3, H4 and H2. While we find the p value for H1 is above the adjusted alpha at the 5% threshold, it is still under the adjusted alpha at the 10% threshold.

Appendix F: Experiment 2: Between-Subject Tests

We conduct between subject tests for H1 and H2 in experiment 2, as designated in our pre-registration. Table F1 reports participants whose first decision task was either Control, 3PG(0), or 3PG(10) and the frequency of the selfish decision in that task respectively. We find that there is a significant difference at the 10% level in selfish behaviour between the Control and 3PG(0) (25.2 vs 35.4, X^2 $p=0.083$). However, testing for H2 we find no significant difference in selfish behaviour between 3PG(0) and 3PG(10) (35.4 vs 30.5, X^2 $p=0.420$). Given the high noise in the Mturk environment, the lower observations in the between subject tests, and less power compared to the within subject tests, these results are not surprising. However, we believe it is encouraging that the results are at least directionally consistent with the hypotheses. A larger sample would be needed to test the existence of these effects in a between-subject-design.

Table F1: Between-Subject Design Tests

Treatment	**Selfish Frequency**	**n**
Control	25.2	131
3PG(0)	35.4	113
3PG(10)	30.5	131

Chapter 3: No Intention to Profit, But Still Repugnance: Evidence From Online Experiments.

Ben Grodeck (Monash University)

Erte Xiao (Monash University)

Nina Xue (Monash University)

Abstract: We study whether and why people feel repugnance towards *harmless* transactions that profit off others' misfortune, without causing the misfortune. Examples include second-hand markets for life insurance, and prediction markets for natural disasters. Repugnance in these contexts can be a constraint on market efficiency. In a series of online experiments (N>2000) that vary in the moral intensity of misfortune—from monetary losses in a game to deaths from road accidents—we find robust evidence of repugnance, measured using costly third-party punishment, towards the party who profits from others' negative outcomes (that are merely determined by luck). Intentions to profit from others' misfortune play a limited role in the punishment decisions. Repugnance is observed even when the profits are associated with good outcomes. Overall, repugnance is mainly outcome-based: people dislike profit-making that occurs as a result of others' (mis)fortune.

JEL codes:

Keywords: repugnance; market; intention; profit; punishment; online experiment

Acknowledgments: Macquarie seminar, SIN Workshop 2023, AYEW 2023, BREW ESA 2022, NTU seminar, Shandon University, 36th PhD Conference in Economics and Business (UWA), NoBEC, NTU.

1. Introduction

Public attitudes towards market transactions are important for the function of markets. Third parties may find certain market transactions to be distasteful, unfair, inappropriate and can sometime even find them repugnant. Transactions perceived as repugnant often face significant constraints and can even be legally banned (Roth, 2007; Elias et al., 2017). Our understanding of the underlying drivers of the repugnant feelings, however, remains limited. One frequently cited explanation is the potential harm imposed by the profiting activities on others, such as exploitation and risk to the buyers or sellers (Satz, 2010; Sandel, 2012; Leuker et al., 2021). Yet, it is unclear whether harm is a necessary condition for repugnance.

We investigate a class of transactions that have a common theme in that they are absent of any direct harm. Specifically, transactions that involve profiting from the misfortune of others even when the transaction does not *cause* the misfortune. Examples include viatical settlements or viagers, in which investors can purchase the life insurance or the property of another party for a cash payment and receives the death benefit or the property upon the insured party's death; prediction markets such as the Policy Analysis Market (PAM), a futures market proposed by a US intelligence agency in which investors can profit from correctly predicting future geopolitical events such as regime changes and terrorist attacks (Looney, 2004),[49] buying foreclosed homes, in which the inability of a previous homeowner to pay their mortgage often leads to the property being sold at a discount.[50] Transactions in these markets have certain benefits. For example, viagers allow elderly homeowners to obtain a regular source of income until their death (The Economist, 2021). Similar to other prediction markets, the prices of the trading contract in PAM can be used to improve the accuracy of forecasts (Hanson, 2007; Arrow et al., 2008; Tetlock & Gardner, 2016). Yet, ethical concerns are often raised about these transactions, and repugnance is a major obstacle in the establishment and the functioning of these markets.

While these transactions do not cause the negative events, we note that a common feature is that profits are attached to the occurrence of others' misfortunes. We call this phenomenon "*piggyback profiting*", whereby one party simply profits from the outcomes of

[49] PAM was widely denounced by politicians as immoral and as "an incentive to commit acts of terrorism". The proposed program was subsequently cancelled by the Pentagon (Schoen, 2003).

[50] Other examples include films that are made about real life tragedies (Burke, 2022; Bushby & Youngs, 2022); short selling that profits from the failure of a company by betting that the price of the company's stocks will fall (for single investors, it is highly unlikely that taking a short position would affect the stock price in a tangible way) (Kelly & Goldstein, 2021).

another, without having any role in the outcome itself.[51] This paper takes a first step to investigate whether people feel repugnance towards harmless piggyback profiting transactions. We further examine potential factors that could explain why people might find such transactions repugnant. Specifically, we test whether repugnance might be driven by the intention to profit from another's misfortune ("*intention-based repugnance*"), or simply the fact that one party's profit is attached to another's bad outcome without any intention to piggyback profit ("*outcome-based repugnance*"). Understanding the role of intentions can have important policy implications for the design of market institutions; we discuss this at the end of the paper. We also explore whether repugnance extends to piggyback profiting from others' *good* fortunes. If so, then similar market constraints could also be present in piggyback profiting transactions based on other's good outcomes. For example, in prediction markets, traders can bet on good outcomes occurring for others and this could also generate repugnance.

We conduct three studies to address our research goals. In Study 1, we design an experimental paradigm in which luck determines the outcome of one player (Player B) in a Rock-Paper-Scissors (RPS) game that is played against the computer. If Player B wins the RPS game (a good outcome), they receive additional earnings; but if B loses the RPS game (a bad outcome), they lose their endowment. A second player (Player A) can piggyback profit from B's outcome. Our main treatment variation is whether Player A has the intention to piggyback profit from Player B's outcome. In the *Bet* treatment, A chooses whether to place a bet on B having a good or a bad outcome. If A places a bet and wins the bet, they can earn a bonus. If A does not bet, their earnings will not depend on B's outcome in the RPS game. In the *Control* treatment, the possibility of piggyback profiting remains but we remove the intention by randomly assigning A to one of three portfolios, in which the earnings correspond to each of the three possible choices by A in the *Bet* treatment. In each treatment, we investigate whether Player B dislikes the piggyback profiting by allowing B to impose costly punishment on A. An advantage of this experimental paradigm is that A's profit directly depends on B's outcome without the possibility that A *causes* B's outcome, good or bad. We can also systematically compare situations in which A's profits are linked to B's good, as opposed to bad, outcomes. Importantly, the comparisons between *Bet* and *Control* inform the role of intention in repugnance, if any.

[51] While we focus on piggyback profiting from others' misfortune, piggyback profiting also occurs from others' good outcomes.

Overall, we find that between 14.7% to 29.9% of Bs choose to punish when A places a potentially profitable bet, confirming the existence of repugnance to piggyback profiting. Interestingly, repugnance is not limited to piggyback profits from others' bad outcomes, as punishment is also observed when A bets that B will have a good outcome. However, punishment frequency is significantly higher for bad outcomes compared to good outcomes. The comparison between the *Bet* and the *Control* treatments shows that intentions to profit play a limited role in punishment decisions. Punishment is significantly more likely to be enacted in *Bet* than *Control*, but only when profits are derived from Player B's bad outcome. We also do not observe significant treatment differences in punishment at the intensive margin.

As repugnance is often associated with outsiders' (or public) attitudes towards the transactions, we conduct Study 2 to check the extent to which the findings in Study 1 can be extended to *unaffected* third parties.[52] Building on the experimental paradigm in Study 1, we introduce an unaffected third party (Player C), whose role is to make the same punishment decision that was made by Player B in Study 1. Again, we find evidence of costly punishment towards piggyback profits. Furthermore, we do not observe intention-based punishment and find similar rates of punishment across *Bet* and *Control*, both when A profits from B's bad outcomes as well as when A profits from B's good outcomes. These results suggest that intentions to piggyback profit do not matter for unaffected third parties. Rather, the association between A's profit and B's outcome (either good or bad) is the main driver of repugnance.

One possible reason why intentions do not matter, especially for unaffected third parties, is that the context of the RPS game is not morally charged. Profiting from someone's outcome in the RPS game is very different from betting on death in a viatical settlement or on the occurrence of a terrorist attack in a prediction market. We thus conduct Study 3, in which instead of betting on outcomes in the RPS game, subjects can bet on whether there will be zero deaths or more than zero deaths on the roads in South Carolina the following day. We observe similar rates of punishment in Study 3 as Study 2, confirming the existence of repugnance towards piggyback profits in a more morally controversial domain. Again, we find that punishment does not increase when the profiting party intentionally makes a bet, rejecting the hypothesis that repugnance is intention-based.

[52] In Study 1, we call Player B an *affected* third party. This is because while they are not involved in the transaction, it is their outcome that is being piggyback profited from.

Our research connects to several strands of the literature. First, there is a growing body of work in economics on repugnant transactions (Roth, 2007; Leider & Roth, 2010; Elias et al., 2017, 2019; Holz et al.(2021) . Repugnance can act as a constraint on market efficiency and yet is hard to predict, as certain transactions are considered repugnant in some cultures but not in others or no longer evoke repugnance as societies evolve (Roth, 2007; Roth & Wang, 2020). Much of this literature has focused on transactions that involve a trade-off between sacred values and secular values (e.g., money), such as the marketisation of human organs and reproductive functions. In addition, previous studies focus on contexts in which the transactions cause harm/benefits at least for someone, such as monetary incentives that distort information acquisition and subsequent behaviour (Ambuehl et al., 2015; Ambuehl, 2017; Stüber, 2021; Kübler & Erkut, 2022). We extend this literature by showing that harmless piggyback profiting that does not have a causal impact on tangible welfare (since the outcome would have occurred even without the transaction) can also lead to repugnance.

Second, our work relates to the large literature in economics and psychology on moral judgments, blame, and reciprocity. One key question is how people perceive (un)kindness and determine whether an action is blameworthy. Theories in this literature often require an agent to cause harm to another and/or intend to harm another even if this harm does not eventuate (Alicke, 1992; Kahneman et al., 1986; Cushman, 2008; Gurdal et al., 2013; Çelen et al., 2017). A number of other studies, however, argue that moral judgment does not necessarily require causality or intentions to harm. Individuals may be deemed morally blameworthy if their intentions are diagnostic about their underlying moral character and the possibility of future harm (Rabin, 1993; Knobe, 2003; Helzer & Pizarro, 2011; Pizarro et al., 2012; Pizarro & Tannenbaum, 2012; Gromet et al., 2016). Closely related to our research, Inbar et al. (2012) argue that actions can be perceived as blameworthy even when they neither cause any harm nor are performed with harmful intentions.[53] Based on the findings from a series of vignettes, the authors propose that moral blameworthiness is driven by the perception of "wicked desires", as the agent chooses to place themselves in a position in which they would be "rooting" for the misfortune to occur. To the best of our knowledge, no study has examined whether intentions in the transaction matters. We fill this gap by designing a novel incentivized experimental paradigm that allows us to distinguish between

[53] An example is the behaviour of Greg Lippmann, a trader at Deutsche Bank who advised investors to bet on mortgage defaults (Senate, 2011).

intention-based and outcome-based repugnance, both for affected and unaffected third parties. Our finding of outcome-based repugnance in third parties suggests that repugnance can occur even when wicked desires are absent or when there is no indication of bad character.

Lastly, we contribute to existing work on perceptions around profit seeking (Kahneman et al., 1986). There is evidence that profit seeking by firms is generally perceived to be incompatible with societal benefits (Bhattacharjee et al., 2017). However, highlighting the potential societal benefits that are accrued when these markets exist can result in more acceptance of markets (Elias et al., 2022). We extend this literature to a setting where profiting activities clearly cause no harm and show that even such harmless transactions can be distasteful.[54]

The rest of the paper is organized as follows. Section 2 introduces the experimental paradigm we use to examine piggyback profiting (Study 1) followed by a conceptual framework to derive our main hypotheses and our main results. Section 3 describes how Study 2 differs from Study 1 and reports the findings from unaffected third parties in the RPS game. Section 4 presents the experimental design and results from Study 3, which features a more morally charged context. Section 5 discusses the results and concludes.

2. Study 1

2.1 Experimental Design

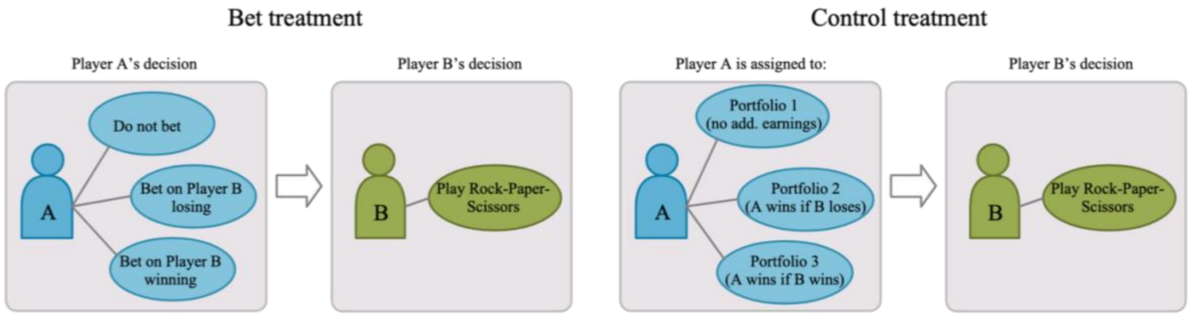
The experiment consists of two tasks and two players, Player A and Player B. Task 1 sets the stage for piggyback profiting. Task 2 elicits attitudes towards Player A profiting off Player B's outcome in Task 1 (see Appendix B for full instructions). Participants receive instructions for both tasks at the beginning of the experiment.

Task 1

[54] Previous research has studied betting behavior in similar settings as ours. Lelieveld et al. (2018) find that audience members of a game show are substantially more likely to bet on a good than a bad outcome for the contestant, which suggests a general reluctance to benefit from others' misfortune. Morewedge et al. (2018) and Kossuth et al., (2020) also find a general reluctance to place bets against outcomes that are personally desirable for the individual (e.g., betting against one's own sports team), suggesting that the act of betting can send a negative signal about one's identity even when betting on one's own misfortune.

In Task 1, both players begin with an endowment of £10. Player B plays a one-shot game of Rock-Paper-Scissors (RPS) against the computer. If Player B has a good outcome in the RPS game, they win an additional £12. However, if they have a bad outcome, they lose their endowment of £10, leaving them with a final payoff of £0.[55] Thus, there is either a good or bad outcome for Player B which depends on random chance, and is outside of Player A's control. We conduct two treatments that vary Player A's role in Task 1 (see Figure 1 below).

Figure 1: Player A and B's decisions in Task 1



***Bet* treatment.** In the *Bet* treatment, Player A can choose whether to place a bet on Player B having a good or bad outcome in the RPS game. Player A makes this decision before Player B plays the game. We chose to make betting costless to ensure that the decision of whether or not to bet is mainly driven by the intention to piggyback profit, rather than other factors such as risk preferences.

If Player A wins the bet, they receive additional earnings of either £0, £4, £8, or £12. Participants are informed that one of these four additional earnings amounts will be randomly selected by the computer. Using a strategy elicitation for each amount, Player A can choose whether they would like to bet on Player B having a good outcome, a bad outcome, or choose not to bet. The four profit conditions allow us to explore whether repugnance varies by the profit amount. We chose £12 as the highest additional earnings of the bet, so that Player A's maximum payoff (£22) does not exceed Player B's maximum payoff from having a good outcome in the RPS game (£22). The £0 profit condition allows us to control for other motivations that may drive punishment. We discuss this in more detail in Section 2.3.

***Control* treatment.** In the *Control* treatment, Player A does not make any betting decisions. Instead, Player A is randomly assigned to one of three portfolios that correspond to

[55] If it was a draw, the computer selected one of the two remaining RPS actions at random. This meant Player A had a 50% chance of having a good outcome and a 50% chance of having a bad outcome.

the payoff structures for Player A in the *Bet* treatment. In Portfolio 1, A does not receive any additional earnings, which is equivalent to Player A choosing not to place a bet in *Bet*. In Portfolio 2, A receives additional earnings if B has a bad outcome in the RPS game. In Portfolio 3, A receives additional earnings if B has a good outcome. Thus, Portfolio 2 (3) is payoff equivalent to A choosing to bet on B having a bad (good) outcome. As in the *Bet* treatment, A's additional earnings in Portfolio 2 or 3, can be either £0, £4, £8, or £12.[56] We inform all participants that Player A does not know to which portfolio they are assigned, and they will never find out. This is to ensure that Player A has no reason to have any "wicked desires" about Player B's outcome in the RPS game.

Task 2

In Task 2, both players start with an endowment of £10. Player B has the option to pay £1, to reduce Player A's payoff (between £0-£10) in this task. We use a strategy method to elicit Player B's punishment decision for each possible bet (either that B has a bad or good outcome) and each possible profit amount.[57] Since there are four profit scenarios (£0, £4, £8, £12) and two outcomes (good, bad) there are eight decisions in total. Equivalently in the *Control* treatment, B decides whether to punish Player A if B is assigned to either Portfolio 2 or 3 for each profit scenario. Figure 2 shows an example of the decision screen in the *Bet* treatment. We inform participants that their punishment decision for one of the four additional earnings amounts will be implemented, if Player A placed a bet for that profit scenario in *Bet* (or if they were assigned to Portfolio 2 or 3 in *Control*). If Player A chose not to bet (or they were assigned to Portfolio 1), Player B's decisions in Task 2 will have no impact on A's payoffs.

While other papers measure repugnance by eliciting participants' willingness to pay (WTP) to stop the transacting from occurring (Kübler & Erkut, 2022), we use costly punishment as a conservative measure of repugnance that is not "cheap talk". We made this

[56] Note that participants can be assigned to Portfolio 2 or 3 even when the profit is £0. This allows us to compare punishment with the *Bet* treatment.

[57] Evidence on whether the strategy method affects punishment is mixed. Karakostas et al. (2022) conduct two meta-analyses and find that the strategy method biases anti-social behaviour upwards. However, there is also evidence that suggests punishment is higher in "hot" conditions (direct-response) than "cold" conditions (strategy-method) (Falk et al., 2005; Brandts & Charness, 2011) . Jordan et al., (2016) find that the strategy method does not affect third-party punishment behaviour. To ensure that the strategy method was not causing confusion, or confounding our results, we ran a follow up experiment with only one additional earnings amount (£12) and found no differences in behaviour. Data can be provided upon request.

design decision for a couple of reasons. First, we use the £0 scenario to construct the piggyback profit variable, which is our dependent variable of interest.[58] Furthermore, eliciting costly punishment for the £0 scenario allowed us to measure if negative beliefs led to repugnance (see Appendix G). Using WTP to stop a £0 transaction doesn't make much sense. Second, we wanted task 1 and task 2 to be independent (we only implemented and paid one of the two tasks). Independence of the repugnant action and measure of repugnance (costly punishment) is important to isolate participants' feelings of repugnance. This would not be possible if task 2 was WTP to stop the transaction, as that decision only be implemented if the decisions in task 1 were also enacted. As a result, we chose to use costly punishment, to measure participant's dislike of piggyback profiting.

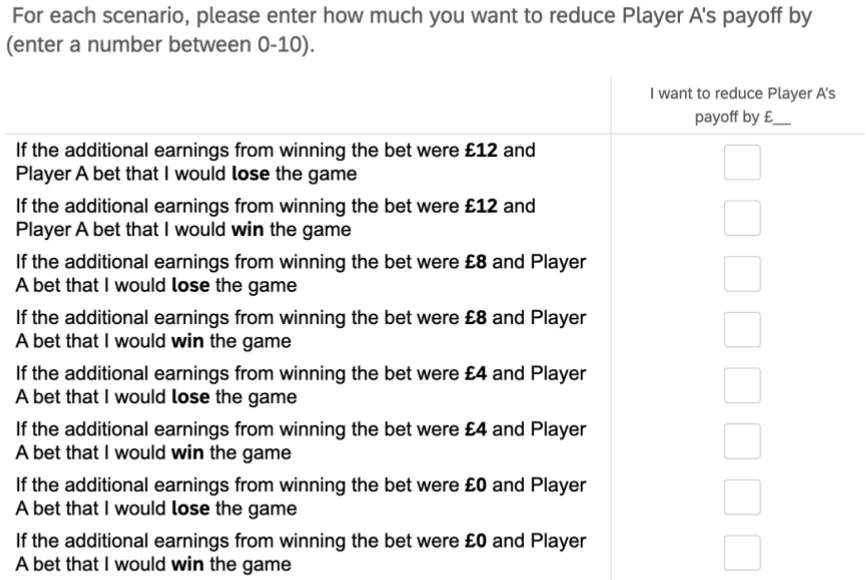
It is important to note that some people who dislike piggyback profiting may not be willing to pay £1 to punish Player A. This would mean that our data measures the lower bound of repugnance. On the other hand, we keep the cost of punishment low so that those who find piggyback profiting distasteful are not discouraged from imposing punishment due to the cost. Moreover, since we are interested in the degree to which people find piggyback profiting distasteful, we set the cost of punishment fixed. If the cost of the punishment increases with the punishment amount, Player B may still impose a small punishment because they are not willing to pay a high cost, even if they find the piggyback profit to be very distasteful.[59]

Player A does not make any decisions in Task 2. However, we do elicit their beliefs about the payoff reduction that Player B will make in each possible scenario. At the end of the experiment, one of the eight scenarios is randomly selected and if Player A guesses Player B's decision correctly, they can receive a bonus of £1.

[58] See section 2.3 for more details.

[59] One might be concerned that Player B will impose the maximum amount of punishment, since the marginal cost of punishment is zero on the intensive margin. However, we do not observe this in the data. In the £12 condition, conditional on punishment only 18.9.% of Player Bs in Bet and 15.9% in Control impose the maximum punishment of £10 if A bet on a bad outcome. Our design may inflate the average punishment amount, but this should not affect the treatment differences or the differences between profit conditions.

Figure 2: Player B's punishment decision



For each scenario, please enter how much you want to reduce Player A's payoff by (enter a number between 0-10).

	I want to reduce Player A's payoff by £__
If the additional earnings from winning the bet were **£12** and Player A bet that I would **lose** the game	
If the additional earnings from winning the bet were **£12** and Player A bet that I would **win** the game	
If the additional earnings from winning the bet were **£8** and Player A bet that I would **lose** the game	
If the additional earnings from winning the bet were **£8** and Player A bet that I would **win** the game	
If the additional earnings from winning the bet were **£4** and Player A bet that I would **lose** the game	
If the additional earnings from winning the bet were **£4** and Player A bet that I would **win** the game	
If the additional earnings from winning the bet were **£0** and Player A bet that I would **lose** the game	
If the additional earnings from winning the bet were **£0** and Player A bet that I would **win** the game	

2.2 Experimental Procedure

The experiment was conducted on the online platform Prolific using Qualtrics software.[60] Before data collection, we used G*Power (Faul et al., 2007) to conduct an a priori power analysis. Our goal was to obtain 0.80 power to detect an effect size of D=0.18-0.28 for our main variables of interest at the standard 0.05 alpha of error probability. We conducted a pilot (N=40 per treatment) for Study 1. Based on the results from the pilot and a priori power calculations, we require approximately n=250 independent pairs (500 participants) per treatment (we recruited 251 pairs in *Bet* and 252 pairs in *Control*). Participants are all residents of the United States.

Each Player A was randomly and uniquely linked to a Player B. Before making any decisions, we asked participants to answer comprehension questions to ensure they understood each task. We first recruited all Player As, followed by the recruitment of Player Bs (this was known to all participants). This meant that Player A's betting decision and

[60] The use of online platforms such as Mturk and Prolific has already received significant uptake in experimental/behavioural economics (Hauser et al., 2014; Exley, 2019; Saccardo & Serra-Garcia, 2020). Gupta et al. (2021) argue that due to the cheaper cost per subject, Prolific provides more power per dollar than the lab. Also, Gandullia et al. (2020) argue, moving from a university student sample to an online sample may also reduce experimenter demand effects as the experimenters are not physically present at the time of data collection, thus further making plausible this choice of participant recruitment.

portfolio assignment always happened before B's punishment decision, to ensure their punishment decisions could not impact A's decisions.

After participants completed both tasks, they were asked to answer some post-survey questions including demographic questions on gender, age, income, political orientation, and highest level of education obtained. All participants were paid £2 to complete the study. The average time to complete the study was approximately 13 minutes. After all participants completed the experiment, 1 in every 20 pairs were randomly selected for bonus payments and their choices were implemented for either Task 1 or Task 2. The experiment was pre-registered on AsPredicted.org (#67931).

2.3 A conceptual framework and hypotheses

We construct a conceptual framework to elaborate on how we use the experiment to measure piggyback profit driven punishment and distinguish intention-based repugnance from outcome-based repugnance. In this framework, Player A's decision of whether to place a bet depends on the monetary profit from betting, the psychological utility from making a correct bet, and their expectation about whether Player B will punish. Below we focus on Player B's punishment decisions in Task 2. The analysis of Player A's betting decision in the *Bet* treatment is provided in Appendix A.[61]

Bet treatment

Let $s \in \{b, g, n\}$ be A's betting decision where b denotes A betting on B receiving a bad outcome, g denotes A betting on B receiving a good outcome, and n denotes A not placing a bet. In our experiment, A's potential additional earnings are $\pi \in \{0, 4, 8, 12\}$. In *Bet*, B punishes A for placing a bet if they feel repugnance towards piggyback profiting. We denote this piggyback-profit driven punishment as $h(\pi, s)$. By definition, when there is no profit to be made, piggyback-profit driven punishment is zero and does not differ if A bets that B will have a good or bad outcome in the RPS game, $h(0, b) = h(0, g) = 0$. We assume that Player B dislikes Player A profiting off their bad outcome more than when Player A profits off their good outcome, $h(\pi, b) > h(\pi, g)$ when $\pi > 0$.

[61] Player A will place a bet if the benefits of winning outweigh the potential cost of punishment in Task 2. This means that if A believes B will win the RPS game, they will only ever bet on a good outcome, but even if they believe that a bad outcome will occur, they may still choose to bet on a good outcome due to concerns about potential punishment for this action.

We consider some other potential factors which could also drive punishment, such as the joy of punishment or anti-social punishment (Abbink & Sadrieh, 2009; Abbink & Herrmann, 2011), possible experimenter demand effects (Zizzo, 2010), or mistakes. We denote punishment driven by these other factors by ε. We assume ε is orthogonal to piggyback-profit driven punishment and is constant across all the conditions. B's decision to punish thus consists of two components:

$$\mathrm{F}(\pi, s) = h(\pi, s) + \varepsilon$$

In the *Bet* treatment we have:

1) When A bets that B will have a bad outcome:

$$\mathrm{F}(\pi, b) = \begin{cases} h(\pi, b) + \varepsilon, & \text{where } \pi > 0 \\ \varepsilon, & \text{where } \pi = 0 \end{cases}$$

2) When A bets that B will have a good outcome

$$\mathrm{F}(\pi, g) = \begin{cases} h(\pi, g) + \varepsilon, & \text{where } \pi > 0 \\ \varepsilon, & \text{where } \pi = 0 \end{cases}$$

Thus, when A bets that B will have either a bad or a good outcome in the game, we can measure punishment of piggyback profiting $h(\pi, s)$ for each scenario:

$$\mathrm{h}(\pi, b) = \mathrm{F}(\pi > 0, b) - \mathrm{F}(\pi = 0, b)\,, \quad (1)$$

$$\mathrm{h}(\pi, g) = \mathrm{F}(\pi > 0, g) - \mathrm{F}(\pi = 0, g)\,, \quad (2)$$

In addition to testing whether piggyback profiting drives repugnance, i.e. $\mathrm{h}(\pi, s) > 0$, our other main research question is whether repugnance is intention-based or whether it can be driven by the outcome alone. Recent literature on moral judgment argues that even without causing any harm, individuals may be perceived as morally blameworthy if their behaviour reveals their underlying moral character and the possibility of future harm (Rabin, 1993; Helzer & Pizarro, 2011; Pizarro et al., 2012; Pizarro & Tannenbaum, 2012; Gromet et al., 2016); or if the actions shape the agent's wicked desires for the harm to occur (Inbar et al., 2012). All of these moral judgement theories are built on the intentions underlying agents' behaviour. However, the literature has yet to provide any empirical evidence that intention is a necessary condition. For example, Inbar et al., (2012) design vignettes to test the wicked desire hypothesis by comparing contingent profit (profit when harm occurs) and non-contingent profit (profit that is independent of the outcome). As both intentions and outcomes change in the two vignettes, this comparison does not inform the possibility that people's judgment is based on the outcome.

Likewise, in our *Bet* treatment, $\mathrm{h}(\pi, s)$ can be both intention- and outcome-based. When Player A bets on Player B's misfortune, the profit opportunity could result in A having the "wicked desire" for B to have a bad outcome in the RPS game. Player A's desire to profit from someone else's outcome could also signal poor moral character. While the wicked desire account should apply only to the case when A bets that B will have a bad outcome, the moral character account can matter when A bets that B will have either a bad or a good outcome.[62] Another possibility is that people simply do not like the pure attachment of A's profit with B's outcome, which does not require any active profit-seeking decisions by A. This can be interpreted as purely *outcome-based repugnance*. Next, we show how the *Control* treatment can help differentiate between intention and outcome-based repugnance.

Control treatment

Player A cannot reveal any intentions to profit from B's outcome in *Control*. Thus, if Player B chooses to punish A, this could only be driven by outcome-based repugnance. We denote pure outcome-based punishment of piggyback profiting as $\mathrm{h}'(\pi, s')$ to differentiate from $h(\pi, s)$ in the *Bet* treatment, where $s' \in \{b', g', n'\}$. A is assigned to either Portfolio 1 which gives no additional earnings regardless of B's outcome ($s' = n'$), Portfolio 2 which gives additional earnings if B has a bad outcome in the game ($s' = b'$), or Portfolio 3 which gives additional earnings if B has a good outcome ($s' = g'$). Applying the framework as in the *Bet* treatment, B's decision to punish consists of the following two components:

$$\mathrm{F}'(\pi, s') = h'(\pi, s') + \varepsilon$$

Similar to the *Bet* treatment, we denote B's decision to punish when A is assigned to Portfolio 2 as $\mathrm{h}'(\pi, b')$; and B's decision to punish when A is assigned to Portfolio 3 as $\mathrm{h}'(\pi, g')$. Thus, in *Control*, we have:

1) When A is assigned to a portfolio that profits from B's bad outcome

$$\mathrm{F}'(\pi, b') = \begin{cases} h'(\pi, b') + \varepsilon, & \text{where } \pi > 0 \\ \varepsilon, & \text{where } \pi = 0 \end{cases}$$

2) When A is assigned to a portfolio that profits from B's good outcome

[62] Note that intention to bet on B's bad outcome could be a sufficient but not a necessary condition for the wicked desire account. Wicked desire is possible when A simply knows that they could profit from B's bad outcome. In the control treatment, we exclude the possibility of wicked desire by ensuring B that A does not know about the source of the profit.

$$\mathrm{F}'(\pi, g') = \begin{cases} h'(\pi, g') + \varepsilon, & \text{where } \pi > 0 \\ \varepsilon, & \text{where } \pi = 0 \end{cases}$$

Thus, we can measure pure outcome-based punishment of piggyback profiting, $\mathrm{h}'(\pi, s')$, as follows:

$$\mathrm{h}'(\pi, b) = \mathrm{F}'(\pi > 0, b') - \mathrm{F}'(\pi = 0, b'), \qquad (3)$$

$$\mathrm{h}'(\pi, g) = \mathrm{F}'(\pi > 0, g') - \mathrm{F}'(\pi = 0, g'), \qquad (4)$$

Taking all these together, by comparing the punishment behaviour in *Bet* and *Control*, we can determine whether piggyback profiting drives punishment and if so, the role of intentions in punishment decisions. If B dislikes A piggyback profiting on their bad outcome i.e., $h(\pi > 0, b) > 0$, then according to equation (1) above, we should expect that, when A bets that B will have a bad outcome, B is more likely to punish A when the bet is profitable than when there is no profit. Furthermore, if A's intention to piggyback profit matters in B's punishment decision, we should expect $h(\pi, b) > h'(\pi, b')$. According to (1) and (3), this means that the difference between punishment in the profit and no profit conditions in *Bet* should be greater than the difference in punishment between the two corresponding scenarios in *Control* (i.e., when A is assigned to Portfolio 2 with a positive profit and when A is assigned to Portfolio 2 with zero profit). On the other hand, this difference-in-difference should not be significant if punishment is not driven by A's intentions but instead by the pure fact that A's profit is attached to B's bad outcome. We summarize these considerations when profit is associated with B's bad outcome in the following preregistered hypotheses.

Hypothesis 1: Punishment is imposed on piggyback profiting from another's bad outcome:

$$F(\pi > 0, b) - F(\pi = 0, b) > 0$$

Hypothesis 1a (Intention-based punishment): Punishment imposed on piggyback profiting is triggered by the intention to profit from another's bad outcome.

$$F(\pi > 0, b) - F(\pi = 0, b) > F'(\pi > 0, b') - F'(\pi = 0, b') > 0$$

Hypothesis 1b (Outcome-based punishment): Punishment is triggered by the attachment of one's profit to another's bad outcome.

$$F(\pi > 0, b) - F(\pi = 0, b) = F'(\pi > 0, b') - F'(\pi = 0, b') > 0$$

Although our main interest is piggyback profiting from another's misfortune, our experiment allows us to explore whether the punishment of piggyback profiting may even occur when one's profit is linked to others' *good* fortunes. For example, if B views A's intention to profit from their good outcome in the game as a negative signal about A's moral character, we would expect to see $h(\pi > 0, g) > 0$, and that punishment is greater in *Bet* than in *Control*, i.e., $h(\pi > 0, g) > h'(\pi > 0, g')$. Or, if B dislikes the attachment of A's profit to B's outcome—good or bad—B would punish when A is assigned to Portfolio 3 and more so when there is a positive profit than when there is no profit, Thus, we have the following hypotheses regarding punishment of piggyback profiting from B's good outcome.

Hypothesis 2: Punishment is imposed on piggyback profiting from another's good outcome:

$$F(\pi > 0, g) - F(\pi = 0, g) > 0$$

Hypothesis 2a (Intention-based punishment): Punishment is triggered by the intention to make a piggyback profit from another's good outcome.

$$F(\pi > 0, g) - F(\pi = 0, g) > F'(\pi > 0, g) - F'(\pi = 0, g) > 0$$

Hypothesis 2b (Outcome-based punishment): Punishment is triggered by the attachment of one's profit to another's good outcome.

$$F(\pi > 0, g) - F(\pi = 0, g) = F'(\pi > 0, g') - F'(\pi = 0, g') > 0$$

Our hypotheses are built on the comparisons between profitable and non-profitable bets. We are agnostic about how repugnance changes as the profit amount increases. One possibility is that repugnance increases as profits from the bet increase. It is also possible that people only care about whether there is a profit or not and the repugnant feeling is insensitive to the profit amount. Another alternative is that there is a threshold such that repugnance occurs only when the profit reaches a certain level. We explore the relationship between profit amount and punishment in our data.

Discussion of assumptions

In the above discussion, we make a few assumptions about B's punishment decisions. One is that B does not punish betting or gambling per se. While gambling may be viewed as repugnant in some cultures, it is mostly acceptable in the U.S. where our subjects reside

(Etuk et al., 2022). Moreover, as the RPS game itself has the nature of gambling, we argue it is unlikely that B will punish A simply because they find gambling distasteful. Furthermore, A cannot gamble in the control treatment. Thus, any punishment driven by the distaste towards gambling per se would lead to an overestimate of the intention-based repugnance. Yet, we report later that we find only weak evidence for intention-based repugnance in our experiment. B does not care if A derives pleasure from betting per se (e.g., pleasure from making a correct prediction).[63] A violation of this assumption should not affect the calculation of $h(\pi, s)$ in (1) and (2) as any negative feeling towards betting per se would be present in both the positive and zero profit conditions.

For simplicity, we also assume that payoff differences (i.e., inequality) between A and B play a minor role in the B's punishment decision. A violation of this assumption should not affect the test of the intention-based repugnance. The reason is that payoff differences are the same across *Bet* and *Control*. This means observed treatment differences in punishment cannot be attributed to inequality aversion. On the other hand, as we will report in the results section, our data point to outcome-based repugnance as the main driver of punishment, which can be potentially explained by the payoff inequality between A and B. In Study 3, we examine the robustness of our results in a more morally charged context in which A can bet on the possibility of a loss of life. It is reasonable to assume that inequality is unlikely to drive punishment decisions as punishment cannot reduce the "inequality" between monetary profit and a loss of life, or at least not to the same extent as in the pure monetary context. We further discuss the role of inequality in Section 5.

Lastly, we assume that B does not punish A's negative beliefs that B is more likely to have a bad than a good outcome in the RPS game. If this assumption is violated, we should observe punishment even when A earns no profit from the bet $F(\pi = 0, b) > 0$ and that punishment is more likely when A bets on B having a bad outcome than when A bets on B having a good outcome $F(\pi = 0, b) > F(\pi = 0, g)$. We find no evidence that negative beliefs affect punishment decisions (see Appendix G).

2.4 Results

We focus on Player B's punishment decisions to test our main hypothesis of piggyback profit driven punishment and the role of intentions versus outcomes. The results of Player A's

[63] If A derives pleasure from making a correct prediction, they would bet even when profit is £0. We consider this possibility when discussing A's betting behavior in Appendix A.

betting decisions and beliefs about punishment are reported in Appendix C. To test our hypotheses, following previous literature on punishment, we compare both punishment frequency (extensive margin) and the mean punishment amount when punishment occurs (intensive margin).

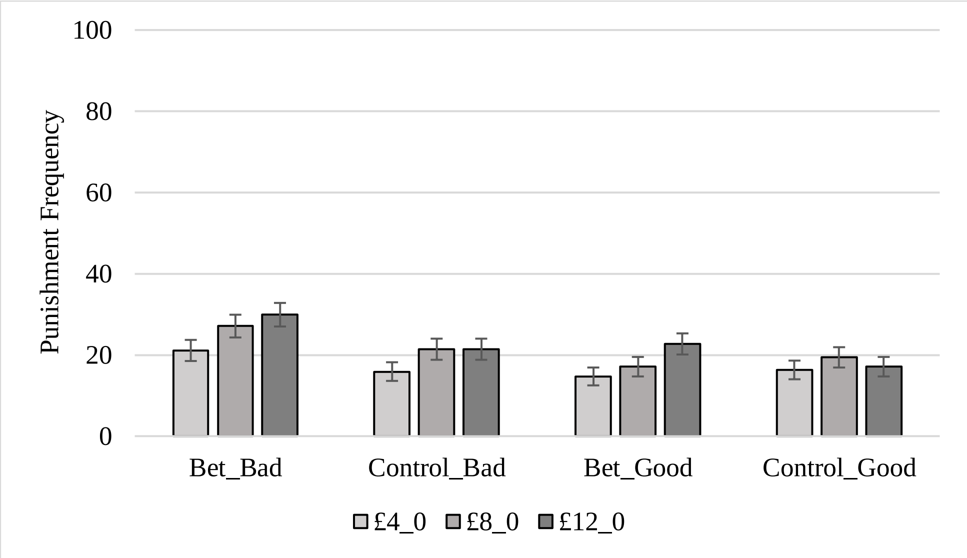
Based on the conceptual framework discussed above, we measure "piggyback profit punishment" as the additional punishment under the positive-profit condition after controlling for punishment that is imposed under the zero-profit condition. For example, suppose Player B imposed punishment of £8 when A bets on B having a bad outcome when the additional profit is £12, and punishment of £2 for betting on a bad outcome when profit is £0. In this case, we define that punishment of piggyback profit has occurred with a punishment amount of £6. Figure 3 reports both the frequency of punishment (extensive margin) and average of punishment amount when piggyback profit punishment occurs (intensive margin), in both treatments for each of the three piggyback profit conditions.

Supporting Hypotheses 1 and 2, in all three profit conditions, we observe that piggyback profit punishment occurs both when A bets on B having a bad outcome and when A bets on B having a good outcome. At the extensive margin (presented in Figure 3a), piggyback profit punishment is significantly greater than 0 in all twelve scenarios (t-test, $p<0.001$). Although punishment occurs both when A bets on B having either a bad or a good outcome, the frequency of piggyback profit punishment is significantly higher for bets on bad outcomes than good outcomes (McNemar's test, £12: 29.9% vs 22.7%, $p=0.002$; £8: 27.1% vs. 17.1%, $p<0.001$; £4: 21.1% vs 14.7%, $p=0.005$).

On the other hand, as shown in Figure 3b, punishment is not significantly different between good and bad outcomes on the intensive margin (Wilcoxon signed-rank test, £12: 4.08 vs 3.89, $p=0.142$; £8: 3.38 vs 3.23, $p=0.314$; £4: 1.66 vs 2.30, $p=0.165$). We also observe that the piggyback profit punishment amount increases as the profit of betting increases. We discuss in more details below that punishment of piggyback profiting on both the extensive and intensive margins significantly decreases, as the profit decreases. However, the relationship between the profit amount and punishment is not linear .

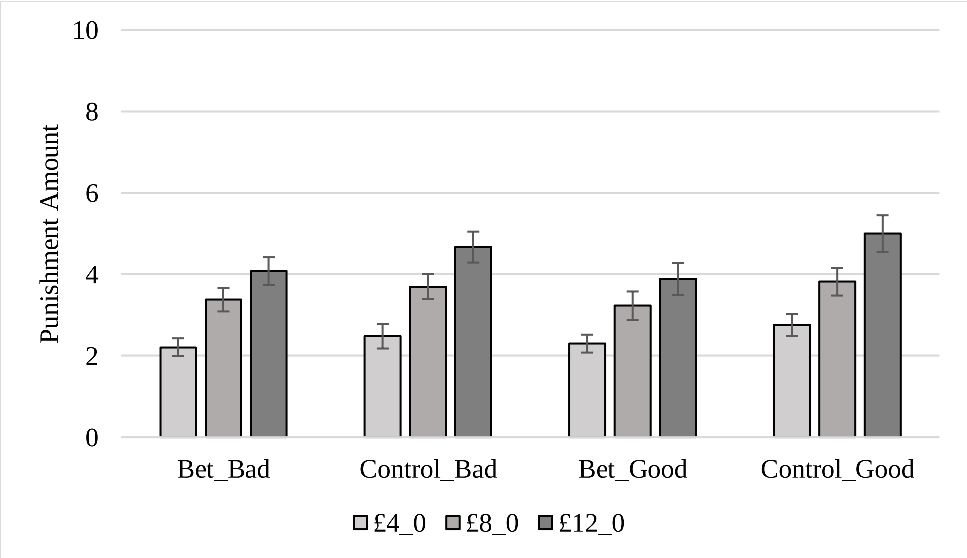
Figure 3. Piggyback profit punishment

(a) Punishment frequency



Note: # observations: Bet_Bad=251; Control_Bad=252; Bet_Good=251; Control_Good=252

(b) Punishment amount



Note: # observations: Bet_Bad: £12=75, £8=68, £4=53; Control_Bad: £12=54, £8=54, £4=40; Bet_Good: £12=57, £8=43, £4=37; Control_Good: £12=43, £8=49, £4=41

To examine the role of intentions in Player B's decision to punish, we compare piggyback profit punishment in the *Bet* and the *Control* treatments. Below we focus on the

£12 profit condition and show that intentions play a limited role in punishment decisions. Results are similar for the other two profit conditions (See Appendix D).

We find that when A can profit from B's bad outcome, piggyback profit punishment is significantly more likely to occur in *Bet* than in *Control*. (£12: 29.9% vs. 21.4%, Chi-squared test, p=0.030). This suggests that intention matters in B's decision on whether to punish. However, the intention does not seem to matter in terms of how much to punish. We observe no significant difference in the average amount of punishment (4.08 vs. 4.67, Wilcoxon rank-sum test, p=0.174). When A can profit from B's good outcome, punishment frequency is directionally higher in the *Bet* treatment, but this difference is not significant (Chi-squared test, 22.7% vs. 17.1%, p=0.113). The average amount of punishment is even significantly lower in *Bet* than in the *Control* (3.89 vs. 5.00, Wilcoxon rank-sum test, p=0.049), an opposite direction of the prediction of intension-based repugnance. This result suggests that intentions are not driving punishment decisions on the intensive margin.

To provide robustness tests for the above results, we further conduct a regression analysis in which we include control variables using the demographic data obtained in the survey, including gender, education, political orientation, religiosity, and income. We use a hurdle model for two key reasons. First, there are a large number of corner solutions in our data, with 70.1% of participants in *Bet* and 78.6% of participants in *Control* not punishing piggyback profiting when profit is £12.[64] In these cases, hurdle models are useful in addressing corner solutions (Cameron & Trivedi, 2010; Wooldridge, 2010). Second, the decision of whether to enact costly punishment is logically separable from the choice of the amount of punishment (Kriss et al., 2016). In our experimental design, punishment is costly on the extensive margin, but not on the intensive margin. We use Cragg's hurdle model (Cragg, 1971) to measure these separate processes. The first hurdle uses a probit regression which examines the binary outcome of enacting punishment or not enacting punishment (step one). Then, if punishment is chosen, the model independently estimates the determinants of the punishment amount using a truncated normal regression (step two).

The hurdle regression results are reported in Table 2. For each regression, the top panel reports the probit regression of the binary decision to enact costly punishment and the bottom panel reports the truncated normal estimates to punish a non-zero amount. The *Control* treatment is the baseline against which the coefficient for *Bet* is compared.

[64] The frequency of participants not punishing is higher in all other scenarios (besides the £12 profit condition for good outcomes in *Bet*, which is 77.3%).

Consistent with the non-parametric tests, we find that the probability of enacting costly punishment is significantly higher in *Bet* than *Control* when A bets on a bad outcome, and no significant difference in the punishment amount (Regressions 1 and 2). We also observe no significant difference in the probability of punishing when A bets on a good outcome (Regressions 3 and 4). For the average punishment amount, the coefficient of Bet is either not significant (Regression 3) or negative when we include demographic data as control variables (Regression 4).

Table 2: Player B's piggyback profit punishment decisions (£12)

VARIABLES	Bad outcome		Good outcome	
	(1)	(2)	(3)	(4)
Probability of Punishing:				
Bet	0.264**	0.319**	0.203	0.240
	(0.122)	(0.135)	(0.128)	(0.145)
Constant	-0.792***	-1.114***	-0.952***	-0.754*
	(0.0886)	(0.354)	(0.0934)	(0.375)
Punishment Amount:				
Bet	-1.057	-1.429	-2.031*	-2.135*
	(0.927)	(0.895)	(1.125)	(1.088)
Constant	3.387***	8.305***	3.807***	12.21***
	(0.872)	(2.231)	(0.979)	(2.659)
Sigma	1.345***	1.213***	1.375***	1.198***
	(0.123)	(0.118)	(0.142)	(0.129)
Controls	N	Y	N	Y
Observations	503	455	503	455

Notes: Cragg hurdle regressions for £12-£0 Profit. Regressions 2 and 4 include Controls. Controls include age, gender, education, political orientation, religiosity, and income. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.[65]

Finally, we also find that profit matters for punishment to a certain degree. We report the regressions in Appendix Tables D2 and D3. In both *Bet* and *Control*, when A profits from B receiving a bad outcome, the frequency of punishment does not significantly increase when profit increase from £8. However, it significantly increases when the profit goes up to

[65] Due to an error in the data for age, the regressions are missing n=48 observations. However, the results are robust when controls are not included.

£12 from £4. In both *Bet* and *Control* we also observe that the average punishment amount significantly increases as profit increase from £8 to £12, and once again from £4 to £8.

When A profits from B receiving a good outcome, punishment frequency in the *Bet* treatment is significantly higher when profit goes up from £8 to £12. The difference between £4 to £8 condition is not significant. In the *Control* treatment, we do not find a significant difference in punishment frequency when piggyback profit is £12, £8, or £4. In both *Bet* and *Control* we once again observe that the average punishment amount significantly increases as piggyback profits increase from £8 to £12, and once again from £4 to £8.

In sum, we find that Player B punishes both piggyback profiting from bad and good outcomes. While we observe some evidence for intention-based punishment, overall, the role of intentions is limited. Intentions only matter for the decision to enact punishment itself, and only when A profits from B's bad outcome. Intentions do not matter on the intensive margin. Next, we describe the design and findings from Study 2, to explore the extent to which these findings apply to unaffected third parties.

3. Study 2 (Unaffected Third party)

The purpose of Study 2 is to investigate whether piggyback profiting is still considered distasteful to a disinterested third party. If we find that these unaffected third parties punish piggyback profiting activities, then this would further support the idea that profiting off others' misfortune is a repugnant transaction.

3.1 Experimental design and procedure

Study 2 follows the same structure as Study 1 with the same two treatments: *Bet* and *Control*. The key difference is that we introduce a third player—Player C—who makes punishment decisions instead of Player B in Task 2.

In Task 1, Player C is an impartial observer. They earn a fixed amount of £22 and do not make any decisions. This payoff structure ensures that Player A never earns more than Player C, therefore inequality aversion between Player C and A cannot be a reason for Player C reducing Player A's payoff.

In Task 2, all three players start with an endowment of £10. In this task, Player C makes the payoff reduction decisions in each scenario (as Player B did in Study 1). Neither

Player A nor Player B make any decisions except guessing the payoff reductions made by Player C. These beliefs are incentivized in the same manner as Study 1.

We follow the same procedures in Study 2 as in Study 1. The study was also implemented on the online platform Prolific using Qualtrics software. We recruited N=783 participants in total: (264 groups in *Bet*; 258 groups in *Control*).[66] All participants are from the USA and had not participated in Study 1. Participants received £2 to complete the study. The average time to complete the study was approximately 13.5 minutes.

3.2. Results

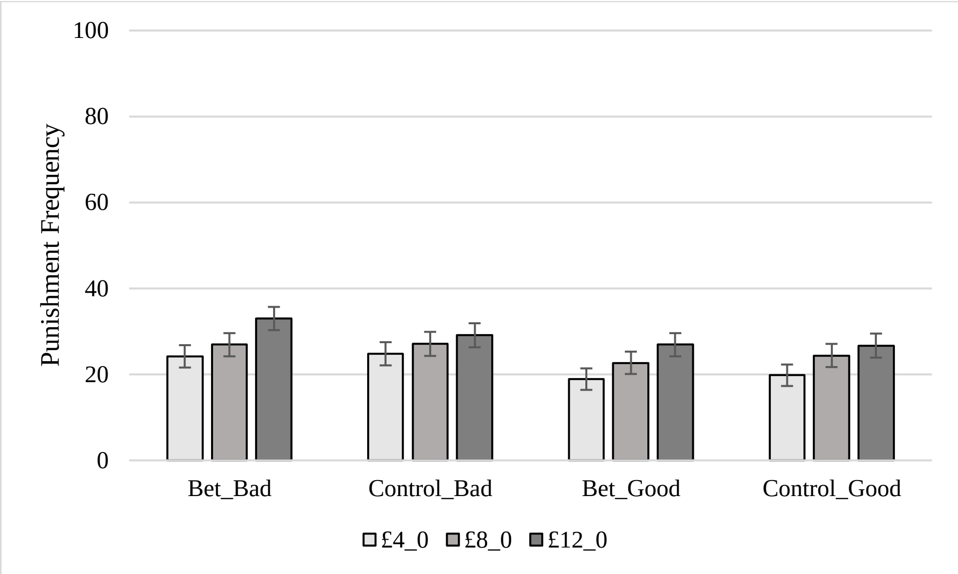
We use the same definitions and conduct the same analyses as in Study 1. Again, we focus on punishment behaviour and report the betting decisions and beliefs of Player A in Appendix E.

Once again, supporting Hypotheses 1 and 2, we observe that piggyback profit punishment occurs both when A bets on B having a bad outcome and when A bets on B having a good outcome. At the extensive margin (presented in Figure 4a) piggyback profit punishment is significantly greater than 0 in all twelve scenarios (t-test, $p<0.001$). These results suggest that piggyback profiting triggers punishment for unaffected third parties. We also observe that piggyback profit punishment on the extensive margin is significantly higher for bets on bad outcomes than good outcomes (£12: 33.0% vs. 26.9%, McNemar's test, $p=0.003$; £8: 26.9% vs 22.7%, McNemar's test, $p=0.052$; £4: 24.2% vs 18.9%, McNemar's test, $p=0.007$), but not on the intensive margin (£12: 4.37 vs 3.94, Wilcoxon signed-rank test, $p=0.214$ £8: 3.60 vs 3.48, Wilcoxon signed-rank test, $p=0.527$; £4: 2.63 vs 2.22, Wilcoxon signed-rank test, $p=0.715$). In Appendix D we report a regression analysis showing that punishment of piggyback profiting on both the extensive and intensive margins significantly increases, as the profit made increases. However, like Study 1, this increase is not linear in the profit amount.

[66] Due to an error in recruitment, we had to recruit 33 extra participants to ensure there were sufficient pairs of A, B and C players in each treatment.

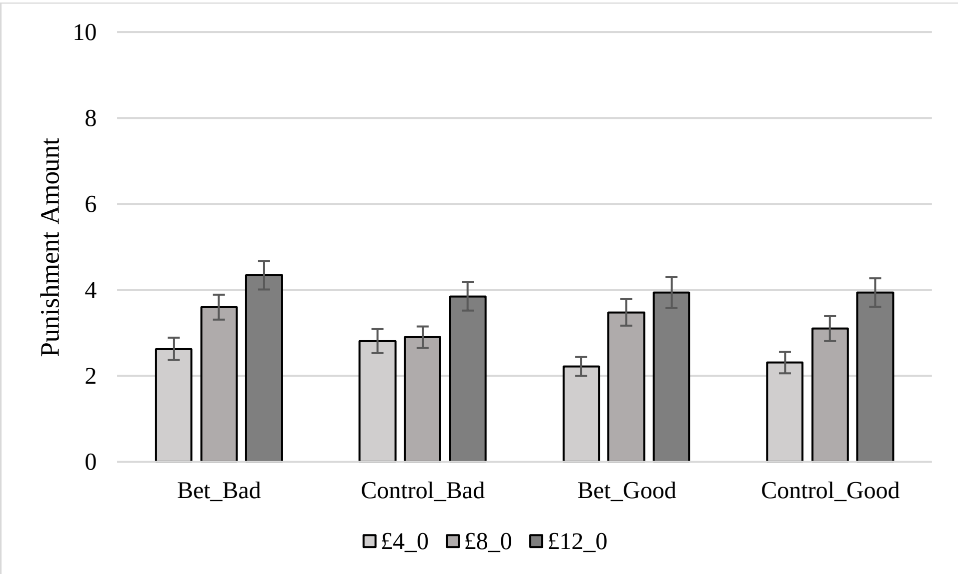
Figure 4: Summary of Piggyback Profit Punishment for Third Party

(a) Punishment frequency



Note: # observations: Bet_Bad=264; Control_Bad=258; Bet_Good=264; Control_Good=258

(b) Punishment amount



Note: # observations: Bet_Bad: £12=87, £8=71, £4=64; Control_Bad: £12=75, £8=70, £4=64; Bet_Good: £12=71, £8=60, £4=50; Control_Good: £12=69, £8=63, £4=51

We once again focus on the £12 piggyback profit case to test intention-based vs outcome-based punishment. To examine the role of intentions in Player C's decision to punish, we conduct the same analysis as in Section 2.4, comparing the difference in the frequency and amount of piggyback profit punishment in the two treatments.

We find that Player C's punishment of piggyback profiting is no different when Player A has an intention to profit from Player B's bad outcome in *Bet* compared to when intentions are not present in *Control* (33.0% vs. 29.1%, Chi-squared test, p=0.337). We also do not find a significant difference between *Bet* and *Control* for the average amount of punishment (4.34 vs.3.85, Wilcoxon rank-sum test, p=0.343). We observe similar results when A can profit from B's good outcome. We find no significant differences between *Bet* and *Control* in Player C's punishment frequency (26.9% vs. 26.7%, Chi-squared test, p=0.969) or average punishment (3.94 vs. 3.94, Wilcoxon rank-sum test, p=0.672). These results suggest that third parties do not engage in intention-based punishment at both the extensive and intensive margins.

In the above analysis, we fail to find a significant difference in punishment frequency between *Bet* and *Control* (no evidence of intention-based punishment). However, we provide further evidence that a null result exists by testing for mean equivalence (Schuirmann, 1987; Dinno, 2017). The mean equivalence test frames the null hypothesis as punishment in the *Bet* treatment being significantly different from the *Control* treatment. The columns represent the equivalence level bounds measured in percentage points. For the Bad Outcome scenario, we observe mean equivalence at the -0.1 to 0.1 threshold. In other words, we can reject the null hypothesis that compared to *Control*, punishment in *Bet* is higher or lower than 10 percentage points. However, we cannot reject the null hypothesis *Bet* increases punishment by more than 7.5 percentage points compared to *Control*. If this treatment effect were to exist, it would still be much smaller compared to the amount of piggyback profit punishment that occurs in *Control*. This means the majority of punishment is still driven by outcome-based punishment. Furthermore, for the £8 and £4 profit scenarios, we observe tighter nulls for bad outcomes at the 7.5 percentage point threshold. We can also reject the null hypothesis that punishment frequency is higher in *Bet* by more than 5 percentage points.[67] These results suggest that if intention-based punishment does exist, it has a minimal impact on participants' decision to

[67] See Appendix D10.

punish.[68] We observe similar results for the good outcome scenario with mean equivalence at the 7.5 percentage point threshold.

Table 3: Mean Equivalence Tests for Punishment Frequency (£12)

	H0: The Bet treatment decreases/increases punishment frequency by x percentage points					
x	-5	5	-7.5	7.5	-10	10
Bet v Control (Bad Outcome)	2.19**	0.28	2.81***	0.89	3.43***	1.51*
Bet v Control (Good Outcome)	1.33*	1.25	1.97**	1.90**	2.62***	2.54***

Note: Z-scores reported above. All t-statistics for TOST procedures on a variety of lower and upper equivalence bounds (in standardized coefficients). *** p<0.01, ** p<0.05, * p<0.1.

We also conduct a regression analysis to provide robustness tests for the above results. Since there is a high proportion of participants who do not punish (in the £12 profit condition for betting on bad outcomes: *Bet*=67.1%, *Control*=70.9%), we once again employ a hurdle model. The Cragg hurdle regression results are reported in Table 4. When piggyback profits are derived from a bad outcome occurring (Regressions 1 and 2), we do not find any evidence that the probability of punishment or the punishment amount is significantly higher in *Bet* than in *Control*. This is also the case for the good outcome scenario (Regressions 3 and 4). This is consistent with the results reported in the non-parametric tests. These results are robust for the profit conditions £8 and £4 (see Appendix D4).

[68] We report the mean equivalence test results for punishment amount in Appendix D11

Table 4: Player C's piggyback profit punishment decisions (£12)

VARIABLES	Bad outcome		Good outcome	
	(1)	(2)	(3)	(4)
Probability of Punishing:				
Bet	0.110	0.148	0.00454	0.0104
	(0.115)	(0.120)	(0.118)	(1.032)
Constant	-0.551***	0.0260	-0.621***	1.680
	(0.0825)	(0.345)	(0.0837)	(3.074)
Punishment Amount:				
Bet	1.179	0.426	0.00392	0.0104
	(1.085)	(0.896)	(1.165)	(1.032)
Constant	0.783	0.126	1.015	1.680
	(1.425)	(2.820)	(1.524)	(3.074)
Sigma	1.494***	1.348***	1.492***	1.387***
	(0.134)	(0.113)	(0.151)	(0.133)
Controls	N	Y	N	Y
Observations	522	522	522	522

Notes: Cragg hurdle regressions for £12-£0 Profit. Regressions 2 and 4 include Controls. Controls include age, gender, education, political orientation, religiosity, and income. Standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05

Finally, we observe that the amount of profit does affect the frequency of punishment (see Appendix Tables D5 and D6). In *Bet*, when A profits from B receiving a bad outcome, the frequency of punishment significantly increases when profit increases from £8 to £12, and from £4 to £8. However, in *Control*, punishment frequency does not significantly increase when piggyback profiting from the bad outcome increases from £8 to £12, but only from £4 to £8. In both *Bet* and *Control* we observe that the average punishment amount significantly increases as profit increases from £8 to £12, but only in *Bet* we also observe a significant increase in average punishment as profit increases from £4 to £8.

We observer a similar patten when A profits from B receiving a good outcome. In *Bet*, the frequency of punishment significantly increases when profits increase from £8 to £12, and from £4 to £8. However, in *Control*, we only observe a significant increase in punishment frequency from £4 to £12. In both *Bet* and *Control* we observe that the average punishment amount significantly increases as profit go from £8 to £12, and £4 to £12 respectively.

Consistent with Study 1, we find that unaffected third parties also punish both piggyback profiting from both bad and good outcomes. However, intentions do not matter for unaffected third parties' punishment decisions on both the extensive and intensive margins. Instead, these results support the outcome-based piggyback profit punishment hypotheses for unaffected third party punishment. One possible explanation for why intentions matter less for unaffected third parties is that the context of the bad outcome (misfortune) is not sufficiently morally charged. We next describe the design and findings from Study 3 which increases the moral stakes of the environment.

4. Study 3 (Moral Context)

We conduct Study 3 by extending Study 2 and explore a misfortune context with greater moral consequences, by replacing betting on monetary outcomes with betting on deaths from road accidents, which captures the nature of the real-life examples of repugnant transactions that involve profiting from death (e.g., viaticals and viagers, prediction markets for natural disasters). This more morally controversial context allows us to examine whether intention-based punishment is more likely to be observed in a more morally charged setting.

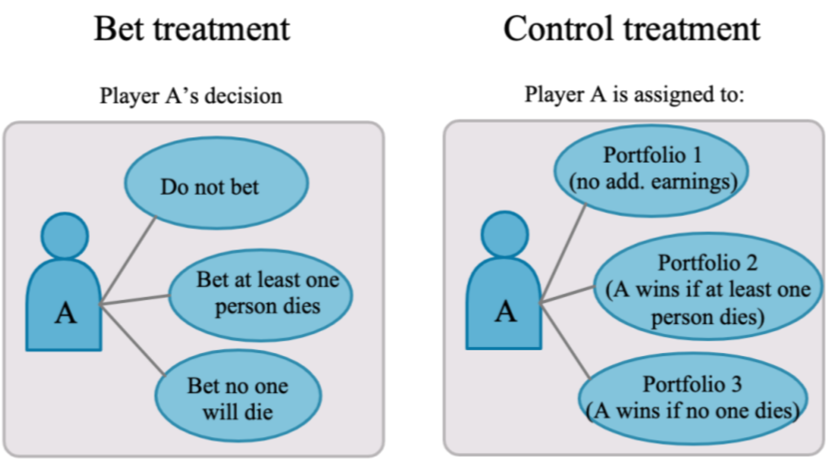
4.1 Experimental design and procedure

Study 3 has a similar structure as Study 2. However, in Task 1, instead of betting on B's outcome in a RPS game, Player A bets on whether there will be more than zero deaths (equivalent to a "bad" outcome) or zero deaths (equivalent to a "good" outcome) on the roads in South Carolina on the next day.[69] Instead of betting on the outcome for a specific

[69] The reason we choose South Carolina is because the South Carolina Department of Public Safety updates the data of fatalities from road accidents daily. The dashboard with data can be found here: https://fatality-count-

participant in the experiment, Player A bets on the outcome for unidentified potential victims. While the stakes are potentially much larger in Study 3, Player A similarly has the option of profiting from others' misfortunes or good fortunes, without playing a causal role in the outcome. Since the misfortune is based on anonymous road death statistics rather than the outcome of a RPS game, the two tasks in the *Bet* treatment involve only Player A, who makes a betting decision in Task 1 and a third party (Player C), who decides whether to punish Player A in Task 2. The incentive structures are kept the same as in Study 2: Player A's endowment is £10 and they receive additional earnings (either £0, £4, £8, £12) if they place a bet and win the bet. Otherwise, their earnings are £10. Player C receives an endowment of £22 in Task 1 irrespective of the road death outcomes. In the *Control* treatment, Player A is randomly allocated to one of three portfolios, which correspond to the three possible choices by Player A in *Bet* (see Figure 5 below).

Figure 5: Player A and B's decisions in Task 1 (Study 3)



Task 2 follows the same procedure as in Study 2, i.e., Player C can enact costly punishment on Player A, if a bet was placed in *Bet* (or if A was randomly assigned to Portfolio 2 or 3 in *Control*). By allowing piggyback profiting from death, the potential function of punishment in reducing the "inequality" between Player A and the victim is

scdps.hub.arcgis.com/apps/fatalities-dashboard/explore. We conducted the experiment on December 8 2021 and the outcome day of interest (road deaths) was December 9 2021. The actual number of fatalities was 2, the average number of deaths per day from the previous week was 4, and the average number of deaths per day in the same month of the previous year was 3. For more information, see the experimental instructions in Appendix B.

arguably weaker as compared to Study 2. In this regard, Study 3 also helps to shed light on the role of inequality aversion in driving the outcome-based punishment observed in Study 2.

We follow the same procedures in Study 3 as in Study 2. Study 3 was also implemented on the online platform Prolific using Qualtrics software. We recruited N=992 participants in total: (254 groups in *Bet*; 242 groups in *Control*). Participants received £2 to complete the study. The average time to complete the study was approximately 11.7 minutes.

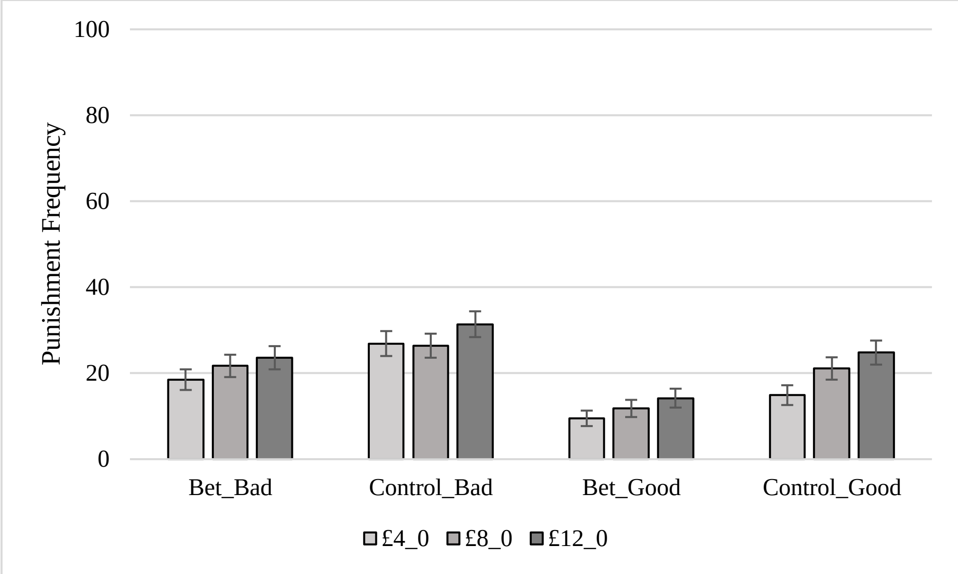
4.2 Study 3 Results

We follow the same procedure as in Study 2 when analysing the data from Study 3. The results of Player A's betting decisions and beliefs about punishment are reported in Appendix F.

Supporting Hypotheses 1 and 2, we once again observe that piggyback profit punishment occurs both when A bets that at least one person will die or when they bet no one will die. At the extensive margin (presented in Figure 3a), piggyback profit punishment is significantly greater than 0 in all twelve scenarios (t-test, $p<0.001$). These results suggest the presence of piggyback profit punishment when there are higher moral stakes. Similarly to the lower moral stakes context, we observe that piggyback profit punishment frequency on the extensive margin is significantly higher for bets on the bad outcome than the good outcome (£12: 23.6% vs 14.2%, McNemar's test, $p<0.001$; £8: 21.7% vs 11.8%, McNemar's test, $p<0.001$; £4: 18.5% vs 9.4%, McNemar's test, $p<0.001$), but not on the intensive margin (£12: 4.02 vs 3.02, Wilcoxon signed-rank test, $p=0.833$ £8: 3.20 vs 2.47, Wilcoxon signed-rank test, $p=0.615$; £4: 2.19 vs 1.71, Wilcoxon signed-rank test, $p=0.157$).

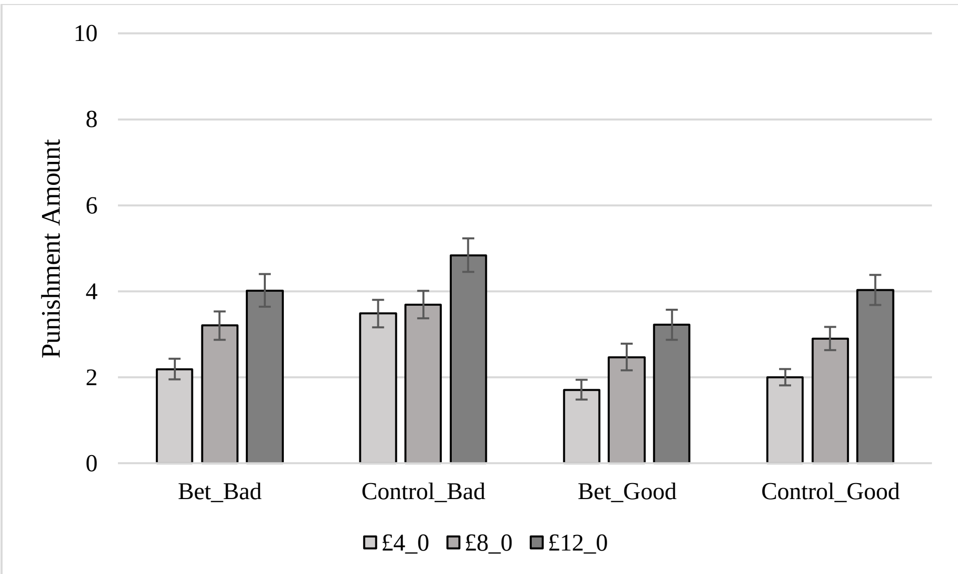
Figure 6. Summary of Piggyback Profit Punishment Moral Context

(a) Punishment Frequency



Note: # observations: Bet_Bad=254; Control_Bad=242; Bet_Good=254; Control_Good=242

(b) Punishment Amount



Note: # observations: Bet_Bad: £12=60, £8=55, £4=47; Control_Bad: £12=76, £8=64, £4=65; Bet_Good: £12=36, £8=30, £4=24; Control_Good: £12=60, £8=51, £4=36

For the testing of intention-based vs. outcome-based repugnance, we focus on the £12 piggyback profit scenario. We conduct the same analysis as in Section 2.4, comparing the difference in the frequency and amount piggyback profit punishment in *Bet* and *Control*.

We find that Player C's punishment of piggyback profiting is directionally less likely when Player A has an intention to profit from a bad outcome (more than 0 deaths) compared to no intention (23.6% vs. 31.4%, Chi-squared test, p=0.052). As punishment in *Control* is higher, this supports the conclusion that intentions do not increase the likelihood that third parties punish.[70] The average punishment amount is also lower in *Bet* than in *Control*, but this difference is not statistically significant (4.02 vs. 4.84, Wilcoxon rank-sum test, p=0.119). Taken together, these results offer evidence that punishment of piggyback profiting on the bad outcome (more people dying) is more outcome-based. We also observe significantly less punishment when A has the intention to piggyback profit from a good outcome (0 deaths) compared to when there is no intention to profit (Chi-squared test, 14.2% vs. 24.8%, p=0.003) We find a directional, but not a significant difference in the punishment amount from a good outcome between *Bet* and *Control* (3.22 vs 4.03, Wilcoxon rank-sum test, p=0.224).

We once again conduct a Cragg hurdle regression analysis as a robustness check for the above results. We report the regression results in Table 5. In Regressions 1 and 2, we observe no significant difference in the probability of punishment or the punishment amount between *Bet* and *Control* when there is a bad outcome (more than 0 deaths). Consistent with the non-parametric results, we observe in Regressions 3 and 4 that participants in *Bet* are significantly less likely to enact costly punishment than in *Control*, but we do not find a significant difference in the punishment amount.

We report the regression analysis for £8 and £4 in Appendix D6. We observe that intentions to profit from a good outcome result in significantly less costly punishment for the £8 profit condition, but not for the £4 profit condition when comparing *Bet* and *Control*. We also observe that both the probability of costly punishment and the punishment amount is significantly less in *Bet* than *Control* for the bad outcome (when the additional profit is £4).

[70] A mean equivalence test shows that we can reject the null hypothesis that punishment frequency is higher in *Bet* than *Control* for bad outcomes by more than 0.1 percentage point (z-score=1.97, p=0.03). For £8, this is 1 percentage point (z-score=1.51, p=0.066) and for £4, 0.1 percentage points (z-score=2.25, p=0.012).

Table 5: Player C's piggyback profit punishment decisions (£12)

VARIABLES	Bad outcome		Good outcome	
	(1)	(2)	(3)	(4)
Probability of Punishment:				
Bet	-0.234	-0.234	-0.392**	-0.401**
	(0.121)	(0.126)	(0.131)	(0.137)
Constant	-0.484***	-0.270	-0.681***	-0.529
	(0.0841)	(0.347)	(0.0877)	(0.376)
Punishment Amount:				
Bet	-1.839	-1.999	-1.609	-1.435
	(1.259)	(1.178)	(1.085)	(0.881)
Constant	2.585*	4.667	2.754**	7.335**
	(1.255)	(3.213)	(0.875)	(2.329)
Sigma	1.543***	1.456***	1.242***	-0.401**
	(0.141)	(0.128)	(0.149)	(0.137)
Controls	N	Y	N	Y
Observations	496	496	496	496

Notes: Cragg hurdle regressions for £12-£0 Profit. Regressions 2 and 4 include Controls. Controls include age, gender, education, political orientation, religiosity, and income. Standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05

Finally, we also find that profit matters for punishment. We report the regressions in Appendix Tables D7 and D8. In *Bet* when Player A can piggyback profit from a bad outcome, we find that the frequency of punishment does not significantly increase when profit increases from £8 to £12, and it significantly increases only when the profit increases from £4 to £8. However, in *Control* we observe the frequency of punishment increases when profit increases from £8 to £12, but not from £4 to £88. In both *Bet* and *Control* we also observe that the average punishment amount significantly increases as profit increases from £8 to £12, but only in *Bet* does it significantly increase as profits increase from £4 to £8.

When Player A profits from a good outcome in *Bet* we also find that the frequency of punishment does not significantly increase when profits increase from £8 to £12, but it significantly increases only when profits increase from £4 to £12. However, in *Control* we observe the frequency of punishment significantly increases both when profits increase from £8 to £12, and from £4 to £8. In both *Bet* and *Control*, we observe that the average punishment amount significantly increases as profits increase from £8 to £12, and also as profits increase from £4 to £8.

In sum, we once again find that Player C punishes piggyback profiting from both a bad and good outcome when the moral stakes are high. We do not observe any evidence in favour of intention-based punishment when Player A can profit from either a bad outcome (more than zero road deaths) or a good outcome (zero road deaths). Interestingly, we find some evidence that intentions may matter when profiting from a good outcome, in the sense that the good intentions result in a lower likelihood of punishment.[71]

5. Discussion

We design a simple experimental paradigm to investigate repugnant attitudes (of both affected and unaffected third parties) towards harmless piggyback profiting from others' misfortune and good fortune. Such repugnance is measured by the difference in the punishment behaviour when the profit is positive and when profit is zero. By examining whether punishment differs depending on whether there is an intention to profit, we further shed light on potential drivers of repugnant attitudes. We find robust evidence that harmless piggyback profiting from others' outcomes—good or bad—is repugnant to both types of third parties. Our results show that intentions play a limited role overall and is only a significant factor in affected third parties' punishment decisions when someone bets on their bad outcome. For third parties, punishment decisions on both the extensive and the intensive margins do not depend on the intention to piggyback profit. This pure outcome-based repugnance by third parties is observed even when profits are associated with a good outcome in a more morally charged context. While previous work has focused more on repugnance towards profiting from others' misfortunes, to the best of our knowledge, we are the first to demonstrate that repugnance extends to profiting from others' *good* fortunes.

The existence of outcome-based repugnance suggests that previous moral judgment theories such as the wicked desires (Inbar et al., 2012) or moral character accounts (Pizarro & Helzer, 2010; Pizarro & Tannenbaum, 2012) cannot fully explain what drives repugnance.

[71] We find some evidence from survey responses that betting on the good outcome was seen in a positive light, e.g., "I deducted half of the bonus earnings in each scenario in which someone died. People should not profit from death. I allowed the full bonus on all bets in which no one dies, as that seems like a victory for everyone!".

Instead, people simply find the concept of making money from someone else's outcomes repugnant. One possibility is that the outcome-based repugnance is driven by inequality-aversion. However, we think inequality aversion is unlikely to be the major reason for the repugnance for two main reasons. First, inequality is held constant between *Bet* and *Control*, so this does not affect the main treatment comparisons. Second, as discussed previously, inequality aversion should play less of a role for punishment decisions in Study 3 when the profit is associated with loss of life while punishment only changes the monetary payoffs. Yet, we observe similar patterns of punishment as in Study 2.

A robust finding within the psychology literature is that agents who act immorally are considered more morally blameworthy when they acted intentionally, meaning they had agency and acted with accurate beliefs and desires about the consequences of the actions (Cushman, 2008; Malle et al., 2014). However, evidence shows that while intentions are important for moral judgements in the harm domain (such as assault), this is not the case for moral judgements in the purity domain (such as consuming taboo substances) (Young & Saxe, 2011; Chakroff et al., 2016). Furthermore, people find purity violations to be more disgusting than harmful acts (Haidt et al., 1993; Royzman et al., 2009) and this correlates with moral judgements. Given that Player A does not cause any harm in our experimental setting, one may plausibly argue that the violation occurs in the purity domain. Our results are consistent with this interpretation since we find that piggyback punishment frequency or punishment amount does not differ between *Bet* and *Control*.

Our finding of outcome-based repugnance has important policy implications for the design of market institutions. The result that intention is not a necessary condition for repugnance means people might find piggyback profiting activities repugnant even if the profiting party is not aware of their involvement in the piggyback profit activity. For example, when the managers of the pension funds invest in viatical settlements, the customers are often not aware of the resource of their profit (Siedle, 2020). Interventions to mitigate the intention to make piggyback profits are unlikely to be effective since the profit is still attached to others' misfortune. More effective policy interventions may include either reducing or eliminating profit opportunities entirely. For example, rather than having private for-profit organisations running betting markets for policy relevant outcomes (such as terrorist attacks, inflation, and natural disasters), governments can promote non-profit prediction platforms (such as Metaculus) or only entrust not-for-profit organisations with the task of managing such markets.

A promising avenue for future research is to investigate whether and how repugnance may be mitigated in order to improve market efficiency. Our experimental paradigm could be adapted to test information interventions aimed at highlighting the fact that these transactions do not actually cause any harm, as well as the societal benefits of allowing piggyback profiting to occur. Other interventions that aim to mitigate repugnance could involve profit-sharing with victims of the misfortune, or having a percentage of profits donated to relevant charities.

References:

Abbink, K., & Herrmann, B. (2011). THE MORAL COSTS OF NASTINESS. *Economic Inquiry*, *49*(2), 631–633. https://doi.org/10.1111/j.1465-7295.2010.00309.x

Abbink, K., & Sadrieh, A. (2009). The pleasure of being nasty. *Economics Letters*, *105*(3), 306–308.

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology*, *63*(3), 368.

Ambuehl, S. (2017). An offer you can't refuse? Incentives change how we inform ourselves and what we believe. *CESifo Working Papers*, *6296*.

Ambuehl, S., Niederle, M., & Roth, A. E. (2015). More money, more problems? Can high pay be coercive and repugnant? *American Economic Review*, *105*(5), 357–360.

Arrow, K. J., Forsythe, R., Gorham, M., Hahn, R., Hanson, R., Ledyard, J. O., Levmore, S., Litan, R., Milgrom, P., & Nelson, F. D. (2008). The promise of prediction markets. In *Science* (Vol. 320, Issue 5878, pp. 877–878). American Association for the Advancement of Science.

Bhattacharjee, A., Dana, J., & Baron, J. (2017). Anti-profit beliefs: How people neglect the societal benefits of profit. *Journal of Personality and Social Psychology*, *113*(5), 671–696. https://doi.org/10.1037/pspa0000093

Brandts, J., & Charness, G. (2011). The strategy versus the direct-response method: A first survey of experimental comparisons. *Experimental Economics*, *14*(3), 375–398. https://doi.org/10.1007/s10683-011-9272-x

Burke, K. (2022, August 4). From Nitram to The Stranger: Daniel Morcombe film reignites debate about 'profiting from pain' | Australian film | The Guardian. *The Guardian*. https://www.theguardian.com/film/2022/aug/04/from-nitram-to-the-stranger-daniel-morcombe-film-reignites-debate-about-profiting-from-pain

Bushby, H., & Youngs, I. (2022, September 30). Netflix's Jeffrey Dahmer drama attracts huge ratings and strong reactions. *BBC News*. https://www.bbc.com/news/entertainment-arts-63088009

Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics using stata* (Vol. 2). Stata press College Station, TX.

Çelen, B., Schotter, A., & Blanco, M. (2017). On blame and reciprocity: Theory and experiments. *Journal of Economic Theory*, *169*, 62–92.

Chakroff, A., Dungan, J., Koster-Hale, J., Brown, A., Saxe, R., & Young, L. (2016). When minds matter for moral judgment: Intent information is neurally encoded for harmful but not impure acts. *Social Cognitive and Affective Neuroscience*, *11*(3), 476–484. https://doi.org/10.1093/scan/nsv131

Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica: Journal of the Econometric Society*, 829–844.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*(2), 353–380.

Dinno, A. (2017). Tost: Two one-sided tests for equivalence. *Stata Software Package*.

Elias, J., Lacetera, N., & Macis, M. (2019). Paying for kidneys? A randomized survey and choice experiment. *American Economic Review*, *109*(8), 2855–2888.

Elias, J., Lacetera, N., & Macis, M. (2022). *Is the Price Right? The Role of Morals, Ideology, and Tradeoff Thinking in Explaining Reactions to Price Surges* (No. w29963; p. w29963). National Bureau of Economic Research. https://doi.org/10.3386/w29963

Elias, J., Lacetera, N., Macis, M., & Salardi, P. (2017). Economic Development and the Regulation of Morally Contentious Activities. *American Economic Review*, *107*(5), 76–80. https://doi.org/10.1257/aer.p20171098

Etuk, R., Xu, T., Abarbanel, B., Potenza, M. N., & Kraus, S. W. (2022). Sports betting around the world: A systematic review. *Journal of Behavioral Addictions*, *11*(3), 689–715.

Exley, C. L. (2019). Using Charity Performance Metrics as an Excuse Not to Give. *Management Science*, mnsc.2018.3268. https://doi.org/10.1287/mnsc.2018.3268

Falk, A., Fehr, E., & Fischbacher, U. (2005). Driving Forces Behind Informal Sanctions. *Econometrica*, *73*(6), 2017–2030. https://doi.org/10.1111/j.1468-0262.2005.00644.x

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Gandullia, L., Lezzi, E., & Parciasepe, P. (2020). Replication with MTurk of the experimental design by Gangadharan, Grossman, Jones & Leister (2018): Charitable giving across donor types. *Journal of Economic Psychology*, *78*, 102268.

Gromet, D. M., Goodwin, G. P., & Goodman, R. A. (2016). Pleasure From Another's Pain: The Influence of a Target's Hedonic States on Attributions of Immorality and Evil. *Personality and Social Psychology Bulletin*, *42*(8), 1077–1091. https://doi.org/10.1177/0146167216651408

Gupta, N., Rigotti, L., & Wilson, A. (2021). The Experimenters' Dilemma: Inferential Preferences over Populations. *ArXiv Preprint ArXiv:2107.05064*.

Gurdal, M. Y., Miller, J. B., & Rustichini, A. (2013). Why blame? *Journal of Political Economy*, *121*(6), 1205–1247.

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, *65*(4), 613.

Hanson, R. (2007). The policy analysis market (a thwarted experiment in the use of prediction markets for public policy). *Innovations: Technology, Governance, Globalization*, *2*(3), 73–88.

Hauser, O. P., Rand, D. G., Peysakhovich, A., & Nowak, M. A. (2014). Cooperating with the future. *Nature*, *511*(7508), 220–223.

Helzer, E. G., & Pizarro, D. A. (2011). Dirty Liberals!: Reminders of Physical Cleanliness Influence Moral and Political Attitudes. *Psychological Science*, *22*(4), 517–522. https://doi.org/10.1177/0956797611402514

Holz, J., Jiménez Durán, R., & Laguna-Müggenburg, E. (2021). Estimating repugnance toward price gouging with incentivized consumer reports. *Available at SSRN 3750332*.

Inbar, Y., Pizarro, D. A., & Cushman, F. (2012). Benefiting From Misfortune: When Harmless Actions Are Judged to Be Morally Blameworthy. *Personality and Social Psychology Bulletin*, *38*(1), 52–62. https://doi.org/10.1177/0146167211430232

Jordan, J., McAuliffe, K., & Rand, D. (2016). The effects of endowment size and strategy method on third party punishment. *Experimental Economics*, *19*(4), 741–763. https://doi.org/10.1007/s10683-015-9466-8

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *The American Economic Review*, 728–741.

Karakostas, A., Tran, N., & Zizzo, D. J. (2022). *Experimental Insights on Anti-Social Behavior: Two Meta-Analyses*.

Kelly, K., & Goldstein, M. (2021, February 8). Wall Street's Most Reviled Investors Worry About Their Fate—The New York Times. *New York Times*. https://www.nytimes.com/2021/02/08/business/wall-street-short-sellers-game-stop.html

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, *63*(3), 190–194.

Kossuth, L., Powdthavee, N., Harris, D., & Chater, N. (2020). Does it pay to bet on your favourite to win? Evidence on experienced utility from the 2018 FIFA World Cup experiment. *Journal of Economic Behavior & Organization*, *171*, 35–58.

Kriss, P. H., Weber, R. A., & Xiao, E. (2016). Turning a blind eye, but not the other cheek: On the robustness of costly punishment. *Journal of Economic Behavior & Organization*, *128*, 159–177. https://doi.org/10.1016/j.jebo.2016.05.017

Kübler, D., & Erkut, H. (2022). *Repugnant Transactions: The Role of Agency and Extreme Consequences*.

Leider, S., & Roth, A. E. (2010). Kidneys for sale: Who disapproves, and why? *American Journal of Transplantation*, *10*(5), 1221–1227.

Lelieveld, G.-J., Inbar, Y., & Van Dijk, E. (2018). Explaining reluctance to benefit from others' misfortune. *Journal of Behavioral Decision Making*, *31*(5), 662–672.

Looney, R. E. (2004). DARPA's policy analysis market for intelligence: Outside the box or off the wall? *International Journal of Intelligence and Counterintelligence*, *17*(3), 405–419.

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A Theory of Blame. *Psychological Inquiry*, *25*(2), 147–186. https://doi.org/10.1080/1047840X.2014.877340

Morewedge, C. K., Tang, S., & Larrick, R. P. (2018). Betting Your Favorite to Win: Costly Reluctance to Hedge Desired Outcomes. *Management Science*, *64*(3), 997–1014. https://doi.org/10.1287/mnsc.2016.2656

Pizarro, D. A., & Helzer, E. (2010). Freedom of the will and stubborn moralism. *Free Will and Consciousness: How Might They Work*, 101–120.

Pizarro, D. A., & Tannenbaum, D. (2012). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil.* (pp. 91–108). American Psychological Association. https://doi.org/10.1037/13091-005

Pizarro, D. A., Tannenbaum, D., & Uhlmann, E. (2012). Mindless, Harmless, and Blameworthy. *Psychological Inquiry*, *23*(2), 185–188. https://doi.org/10.1080/1047840X.2012.670100

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review*, 1281–1302.

Roth, A. E. (2007). Repugnance as a Constraint on Markets. *Journal of Economic Perspectives*, *21*(3), 37–58. https://doi.org/10.1257/jep.21.3.37

Roth, A. E., & Wang, S. W. (2020). Popular repugnance contrasts with legal bans on controversial markets. *Proceedings of the National Academy of Sciences*, *117*(33), 19792–19798.

Royzman, E. B., Leeman, R. F., & Baron, J. (2009). Unsentimental ethics: Towards a content-specific account of the moral–conventional distinction. *Cognition*, *112*(1), 159–174. https://doi.org/10.1016/j.cognition.2009.04.004

Saccardo, S., & Serra-Garcia, M. (2020). Cognitive flexibility or moral commitment? Evidence of anticipated belief distortion. *Evidence of Anticipated Belief Distortion (August 18, 2020)*.

Schoen, J. w. (2003, July 30). Pentagon kills "terror futures market." *MSNBC*.

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, *15*(6), 657–680.

Senate, U. S. (2011). *Wall Street and the financial crisis: Anatomy of a financial collapse*.

Siedle, E. (2020, April 8). *Your Pension May Be Gambling On Human Life, Profiting From COVID Deaths*. Forbes. https://www.forbes.com/sites/edwardsiedle/2020/08/04/your-pension-may-be-gambling-on-human-life-profiting-from-covid-deaths/

Stüber, R. (2021). Why High Incentives Cause Repugnance: A Framed Field Experiment. *Available at SSRN 3850618*.

Tetlock, P. E., & Gardner, D. (2016). *Superforecasting: The art and science of prediction*. Random House.

The Economist. (2021, May 29). The pandemic revives interest in a morbid French financial scheme. *The Economist*. https://www.economist.com/finance-and-economics/2021/05/27/the-pandemic-revives-interest-in-a-morbid-french-financial-scheme

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition*, *120*(2), 202–214. https://doi.org/10.1016/j.cognition.2011.04.005

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*(1), 75–98. https://doi.org/10.1007/s10683-009-9230-z

Appendix

Appendix A: Player A's betting decisions in Study 1

Player A decides whether to place a bet, and if they choose to place a bet, whether to bet on Player B winning or losing the game of RPS. Following the notation in the analysis of Player B's behaviour where $s \in \{b, g, n\}$ denotes A's betting decision and A's additional earnings are represented by $\pi \in \{0, 4, 8, 12\}$. Let $w(s)$ represent the psychological utility from winning a bet (e.g., the pleasure of making a correct prediction). $w(b) > 0$ ($w(g) > 0$) if A bets that B will have a bad outcome (good outcome) in the game and B indeed receives a bad (good) outcome. For simplicity, we assume $w(b) = w(g) = w$. That is, there is no difference between winning a bet based on a bad outcome for B and winning a bet based on a good outcome for B. We also assume that the psychological utility of betting is zero if A places a bet that is not correct, or if A does not bet. Let p be A's belief of the probability that B will have a bad outcome in the game. Player A's utility from their betting decision is therefore represented by:

$$U[\pi, w, F(\pi, s)]$$

whereby Player A weighs the benefits of placing a bet (π, w) against potential punishment by Player B, $F(\pi, s) = h(\pi, s) + \varepsilon$. Recall, we assume that Player B dislikes Player A piggyback profiting off their losses more than Player A profiting off their wins, $h(\pi, b) > h(\pi, g)$ when $\pi > 0$.

Player A, therefore, compares the expected utility of betting that B will have a bad outcome (denoted as $EU(b)$), betting that B will have a good outcome (denoted as $EU(g)$) and not betting (denoted as $EU(n)$):

$$EU(b) = p * \pi + \mathrm{p} * \mathrm{w} - h(\pi, b) - \varepsilon$$
$$EU(g) = (1 - p) * \pi + (1 - \mathrm{p}) * \mathrm{w} - h(\pi, g) - \varepsilon$$
$$EU(n) = 0$$

To facilitate the discussion below, we use $u(s)$ to denote the sum of monetary profit and the psychological enjoyment of betting: $u(s) = \pi + w$. Note that the profit from betting is exogenously determined, and A decides whether to bet for each profit condition. Thus, we can rewrite the expected utilities above as:

$$EU(b) = p * u - h(\pi, b) - \varepsilon$$
$$EU(g) = (1 - p) * u - h(\pi, g) - \varepsilon$$
$$EU(n) = 0$$

Case 1: when profit is positive $\boldsymbol{\pi > 0}$

1) When Player A believes that Player B is more likely to have a good than bad outcome in the game ($p \leq 0.5$):

   Player A will never bet that B will have a bad outcome since $EU(g) > EU(b)$ is always satisfied. Specifically, Player A will:

   - bet that Player B will have a good outcome if $EU(g) > 0$:

   $$u > \frac{h(\pi, g) + \varepsilon}{1 - p}$$

   - not bet if $EU(g) < 0$:

   $$u < \frac{h(\pi, g) + \varepsilon}{1 - p}$$

2) When Player A believes that Player B is more likely to have a bad than good outcome in the game ($p > 0.5$):

   Player A may place a bet on a good or bad outcome. Specifically, A will bet that Player B will have as good outcome if $EU(g) > EU(b)$ and $EU(g) > 0$:

   $$u < \frac{h(\pi, b) - h(\pi, g)}{2p - 1}$$

   and

   $$u > \frac{h(\pi, g) + \varepsilon}{1 - p}$$

   Player A will bet that Player B will have a bad outcome if $EU(g) < EU(b)$ and $EU(b) > 0$.

   $$u > \frac{h(\pi, b) - h(\pi, g)}{2p - 1}$$

   and

$$u > \frac{h(\pi, b) + \varepsilon}{p}$$

It is easy to see that when $\frac{h(\pi,g)+\varepsilon}{1-p} < \frac{h(\pi,b)+\varepsilon}{p}$, we have $\frac{h(\pi,b)+\varepsilon}{p} < \frac{h(\pi,b)-h(\pi,g)}{2p-1}$. Likewise, when $\frac{h(\pi,g)+\varepsilon}{1-p} < \frac{h(\pi,b)+\varepsilon}{p}$, we have $\frac{h(\pi,b)-h(\pi,g)}{2p-1} < \frac{h(\pi,b)+\varepsilon}{p}$.

Thus, A's decision can be summarized as follows:

When $\frac{h(\pi,g)+\varepsilon}{1-p} < \frac{h(\pi,b)+\varepsilon}{p} < \frac{h(\pi,b)-h(\pi,g)}{2p-1}$, A will

- not bet when $u < \frac{h(\pi,g)+\varepsilon}{1-p}$;
- bet that B wins when $\frac{h(\pi,g)+\varepsilon}{1-p} < u < \frac{h(\pi,b)-h(\pi,g)}{2p-1}$
- bet that B loses when $u > \frac{h(\pi,b)-h(\pi,g)}{2p-1}$

When $\frac{h(\pi,b)-h(\pi,g)}{2p-1} < \frac{h(\pi,b)+\varepsilon}{p} < \frac{h(\pi,g)+\varepsilon}{1-p} <$, A will

- not bet when $u < \frac{h(\pi,b)+\varepsilon}{p}$;
- bet that B loses when $u > \frac{h(\pi,b)+\varepsilon}{p}$

Case 2: when profit is zero $\pi = 0$

In this case, $u = w$. That is, the benefit of betting is just the pleasure of winning.

$$EU(b) = p * w - \varepsilon$$
$$EU(g) = (1 - p) * w - \varepsilon$$
$$EU(n) = 0$$

- When $p \leq 0.5$, Player A will bet that Player B will have a good outcome if w $> \frac{\varepsilon}{1-p}$, otherwise no bet will be placed.
- When $p > 0.5$, Player A will bet that Player B will have a bad outcome if w $> \frac{\varepsilon}{p}$, otherwise no bet will be placed.

The above analysis has a few implications. First, when Player A can profit from placing a bet, if A bets that B will have a bad outcome, it implies A believes a bad outcome is more likely than a good outcome ($p > 0.5$). However, if A bets that B will have a good outcome, we cannot conclusively say whether $p > 0.5$ or $p < 0.5$. This is because A may hold the belief that B will have a bad outcome, but bet on a good outcome because they expect greater punishment on betting on a bad outcome compared to betting on a good outcome. On the other hand, when Player A cannot profit from placing a bet, we can draw inference on A's belief based on A's betting decisions. If A bets on a good outcome, A believes that B is more likely to have a good outcome than a bad outcome. Likewise, if A bets on a bad outcome, A believes B is more likely to have a bad outcome than a good outcome. However, some Player As who receive utility from being correct (w), may not choose to place a bet that B will have a good outcome if $0 < \mathrm{w} < \frac{\varepsilon}{1-p}$, nor will they choose to place a bet on a bad outcome if $0 < \mathrm{w} < \frac{\varepsilon}{p}$.

Also, note that in our experiment, if A holds a rational belief that $p = 0.5$ and expects punishment to be greater when betting on a bad outcome than a good outcome, then A should never bet on a bad outcome because she is always better off by betting on a good outcome. A may decide not to bet at all if the benefit of betting is less than the potential punishment.

APPENDIX B: Experimental instructions

<u>Introduction:</u>

This survey is approved by Monash University. All information provided below will be truthful including your interaction with other participants.

In this survey, you will participate in two tasks (Task 1 and Task 2) where you may earn an extra payment. At the end of the survey, we will randomly select 1 in every 20 participants. If you are selected, the computer will randomly **select one of the two tasks**. Your earnings in this selected task will be paid to you via the bonus function on Prolific.

<u>Instructions Part 1: Affected Third Party</u>

In this survey half the participants will be assigned to the role of "**Player A**" and the other half to "**Player B**". Each Player A will be randomly, and uniquely paired with a Player B. Please note your identity will remain anonymous, as will all other participants' identities. You will remain in the same role in the two tasks.

<u>Instructions Part 2: Affected Third Party BET Treatment (Study 1)</u>

We will explain both Task 1 and Task 2 before you find out which role you are in and make your decisions.

Task 1:

In this task, both Player A and Player B start with an **endowment of £10**. The final payoff of each player may change depending on the decisions as described below.

Player B

Player B will play one game of Rock-Paper-Scissors against the computer. The computer's decision will be random, with each possible action having an equal chance of occurring. Note: if the game is a draw, the computer will randomly choose one of the two remaining actions to play against the chosen action of Player B. This means Player B has a 50% chance of winning the game and a 50% chance of losing the game.

- If Player B **wins** the game, they will receive an additional £12. Their **final payoff is £22**.
- If Player B **loses** the game, they will lose their endowment of £10. Their **final payoff is £0**.

Player A

Before Player B plays the game, Player A will have the opportunity to place a bet that Player B will win or lose the game of Rock-Paper-Scissors. If Player A places a bet and **wins the bet**, they can receive additional earnings of **either £12, £8, £4, or £0**. The computer will randomly choose one of the four amounts with equal chances. Player A will know the amount of additional earnings **before** deciding whether to place a bet and what outcome to bet on.

Below are the details of the possible earning outcomes.

- If Player A **does NOT place the bet**, their payoff is just the **endowment of £10.**
- If Player A **places the bet that B will lose** the Rock-Paper-Scissors game.
    - If B **loses** the game, Player A **wins** the bet. Player A's payoff is the **endowment of £10 plus additional earnings from winning the bet (either £12, £8, £4, or £0)**.
    - If B **wins** the game, Player A **loses** the bet. Player A's payoff is just the **endowment of £10**.
- If Player A **places the bet that B will win** the Rock-Paper-Scissors game.
    - If B **loses** the game, Player A **loses** the bet. Player A's payoff is just the **endowment of £10**.
    - If B **wins** the game, Player A **wins** the bet. Player A's payoff is the **endowment of £10 plus additional earnings from winning the bet (either £12, £8, £4, or £0).**

Decision procedure:

1. The computer will randomly choose the amount of additional earnings from winning the bet. Player A will decide whether or not to place a bet and if so, whether to bet on Player B winning or losing the Rock-Paper-Scissors game.

2. Player B will play the Rock-Paper-Scissors game.

Below is a flowchart of Task 1

<u>Instructions Part 3: Affected Third Party BET Treatment (Study 1)</u>

Task 2:

In this task, both Player A and Player B start with an **endowment of £10**.

- If Player A **places a bet** in Task 1, Player B has the opportunity to **pay £1 to reduce Player A's payoff in Task 2** based on Player A's decisions. The maximum reduction Player B can impose on Player A is £10.
    - **Note:** If Player B chooses to impose this reduction, the cost of £1 will be taken from B's £10 endowment in this task. The corresponding reduction amount will be deducted from Player A's £10 endowment and will **NOT** go to Player B's earnings.

This reduction will be made as long as Player A has placed a bet in Task 1, **regardless of whether Player A won or lost the bet**. Player B does **NOT** know the outcome of the Rock-Paper-Scissors game when making the reduction decision.

- If Player A **does not place a bet** in Task 1, both players will earn £10 in this task, no matter what reduction decision Player B makes.

Results: At the end of the survey, if you are one of the 1 in 20 randomly selected participants for extra payment, you will receive information about your own final payoff.

<u>Instructions Part 2: Affected Third Party CONTROL Treatment (Study 1)</u>

Task 1:

In this task, both Player A and Player B start with an **endowment of £10**. The final payoff of each player may change depending on the decisions as described below.

Player B

Player B will play one game of Rock-Paper-Scissors against the computer. The computer's decision will be random, with each possible action having an equal chance of occurring. Note: if the game is a draw, the computer will randomly choose one of the two remaining actions to play against the chosen action of Player B. This means Player B has a 50% chance of winning the game and a 50% chance of losing the game.

- If Player B **wins** the game, they will receive an additional £12. Their **final payoff is £22**.
- If Player B **loses** the game, they will lose their endowment of £10. Their **final payoff is £0**.

Player A

Before Player B plays the game, the computer will randomly assign Player A to **one of the three earnings portfolios**. "**Portfolio 1**" does not provide any additional earnings to Player A. In "**Portfolio 2 (A wins if B loses)**", and "**Portfolio 3 (A wins if B wins)**", depending on whether Player B wins or loses the game of Rock-Paper-Scissors, Player A may receive additional earnings of **either, £12, £8, £4, or £0**. The computer will randomly choose one of the four amounts with equal chances.

Below are the details of the possible earning outcomes of each portfolio.

- "**Portfolio 1**": Player A's payoff is just the **endowment of £10**.
- "**Portfolio 2 (A wins if B loses)**":
  - If B **loses** the game of Rock-Paper-Scissors, Player A's payoff is the **endowment of £10 plus additional earnings (either £12, £8, £4, or £0)**.
  - If B **wins** the game, Player A's payoff is just the **endowment of £10**.
- "**Portfolio 3 (A wins if B wins)**":
  - If B **loses** the game of Rock-Paper-Scissors, Player A's payoff is just the **endowment of £10**.
  - If B **wins** the game, Player A's payoff is the **endowment of £10 plus additional earnings (either £12, £8, £4, or £0)**.

Decision procedure:

1. Player A will be randomly assigned to one of the three portfolios.

2. Player B will play the Rock-Paper-Scissors game.

Note: Player A will **NEVER** know which portfolio they are assigned to.

Below is a flowchart of Task 1.

Instructions Part 3: Affected Third Party CONTROL Treatment (Study 1)

Task 2:

In this task, both Player A and Player B start with an **endowment of £10**.

- If Player A is assigned to "**Portfolio 2 (A wins if B loses)**", or "**Portfolio 3 (A wins if B wins)**" in Task 1, Player B has the opportunity to **pay £1 to reduce Player A's payoff in Task 2**. The maximum reduction Player B can impose on Player A is £10.
    - **Note:** If Player B chooses to impose this reduction, the cost of £1 will be taken from B's £10 endowment in this task. The corresponding reduction amount will be deducted from Player A's £10 endowment and will **NOT** go to Player B's earnings.

This reduction will be made as long as Player A was assigned to "**Portfolio 2 (A wins if B loses)**", or "**Portfolio 3 (A wins if B wins)**", in Task 1, **regardless of whether Player A received additional earnings**. Player B does **NOT** know the outcome of the Rock-Paper-Scissors game when making the reduction decision.

- If Player A is assigned to "**Portfolio 1"** in Task 1, both players will earn £10 in this task, no matter what reduction decision Player B makes.

Results: At the end of the survey, if you are one of the 1 in 20 randomly selected participants for extra payment, you will receive information about your own final payoff.

Instructions for Task 1: Moral Context Bet Treatment (Study 3)

In this task, Player A starts with an **endowment of £10**, and Player B starts with an **endowment of £22**. The final payoff of each player may change depending on the decisions as described below.

Player A
Player A will have the opportunity to place a bet whether anyone will die on the roads in South Carolina on Thursday 9th December 2021. This bet will be evaluated using data from the South Carolina Department of Public Safety.

If Player A places a bet and **wins the bet**, they can receive additional earnings of **either £12, £8, £4, or £0**. The computer will randomly choose one of the four amounts with equal chances. Player A will know the amount of additional earnings **before** deciding whether to place a bet and what outcome to bet on.

Below are the details of the possible earning outcomes.

- If Player A **does NOT place the bet**, their payoff is just the **endowment of £10.**
- If Player A **places the bet** that **at least one person** will die on the roads.
  - If **at least one person** dies, Player A **wins** the bet. Player A's payoff is the **endowment of £10 plus additional earnings from winning the bet (either £12, £8, £4, or £0)**.
  - If **no one** dies, Player A **loses** the bet. Player A's payoff is just the **endowment of £10**.
- If Player A **places the bet** that **no one** will die on the roads.
  - If **at least one person** dies, Player A **loses** the bet. Player A's payoff is just the **endowment of £10**.
  - If **no one** dies, Player A **wins** the bet. Player A's payoff is the **endowment of £10 plus additional earnings from winning the bet (either £12, £8, £4, or £0).**

Player B
Player B has no decision to make in Task 1 and their final payoff is their endowment of **£22**.

APPENDIX C: Study 1 Player A Results

C1: Player A's betting decisions

Table C1 reports the betting decisions of Player A in Bet. We find that in all positive profit scenarios, Player A is more likely to bet on B having a good outcome than a bad outcome in the RPS game in the £12 profit condition (binomial test, $p=0.012$), £8 (binomial test, $p<0.001$), and £4 (binomial test, $p=0.001$). However, there is no difference between bets on the good and bad outcome when the profit is £0 (binomial test, $p=0.248$)[72]. Betting is also much more likely when there is a positive profit compared to no profit (£4 v £0: 74.5% vs 29.9%, McNemar's test, $p<0.001$)[73]. However, we do observe about 30% of participants who bet even when there is no profit, which suggests some intrinsic interest in betting per se (i.e., $w > 0$).

We also asked Player A if they believed that Player B would be more likely to have a bad outcome, good outcome, or equally likely to have a good or bad outcome in the game of RPS. The majority of participants (69.4%) believed that a bad outcome was just as likely as a good outcome. For the remaining participants, the proportion that believed a bad outcome was more likely is slightly higher than the proportion that believed a good outcome was more likely, but the difference is not statistically significant (18.1% vs.12.5%, binomial test, $p=0.154$). According to our conceptual framework, these results suggest that A having a preference of betting that B has a good outcome is unlikely to be driven by a biased belief that B is more likely to have a good outcome than a bad outcome in the game of RPS. Instead, Player A may anticipate that Player B will punish piggyback profiting and that punishment is more likely when A bets on a bad outcome than on a good outcome (see Appendix A). Alternatively, they may just prefer to receive additional earnings when a good outcome happens to Player B.

[72] The null hypothesis for the binomial test is that the probability of betting on either a good or a bad outcome is $p=0.5$. This is the null we use for all binomial tests unless otherwise specified.
[73] We also find $p<0.001$ for £12 v £0, and £8 v £0.

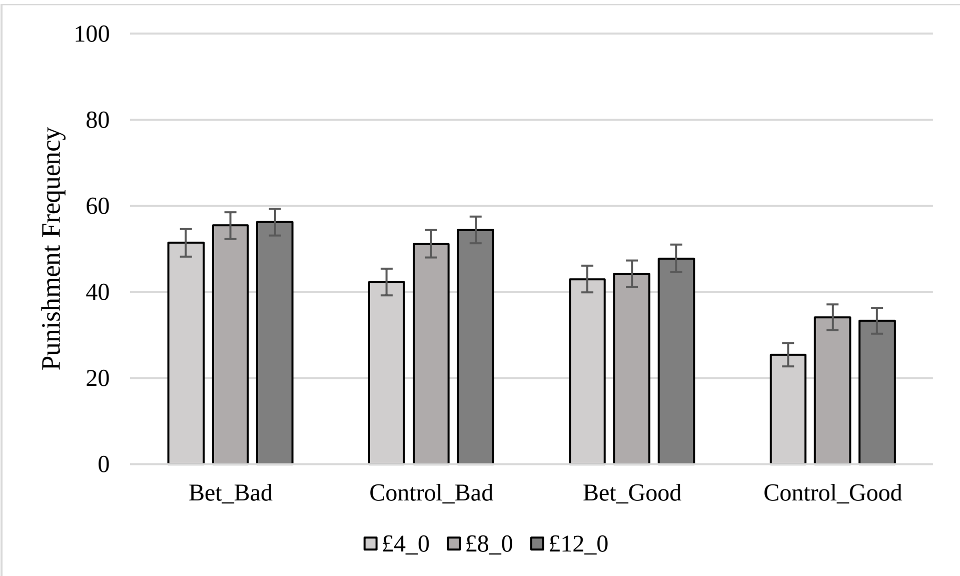
Table C1: Betting decisions of Player A (Bet Treatment)

Profit	Bet on bad outcome	Bet on good outcome	No bet
£12	35.5%	50.6%	13.9%
£8	29.5%	53.0%	17.5%
£4	28.3%	46.2%	25.5%
£0	12.8%	17.1%	70.1%

C2: Player A's beliefs about punishment

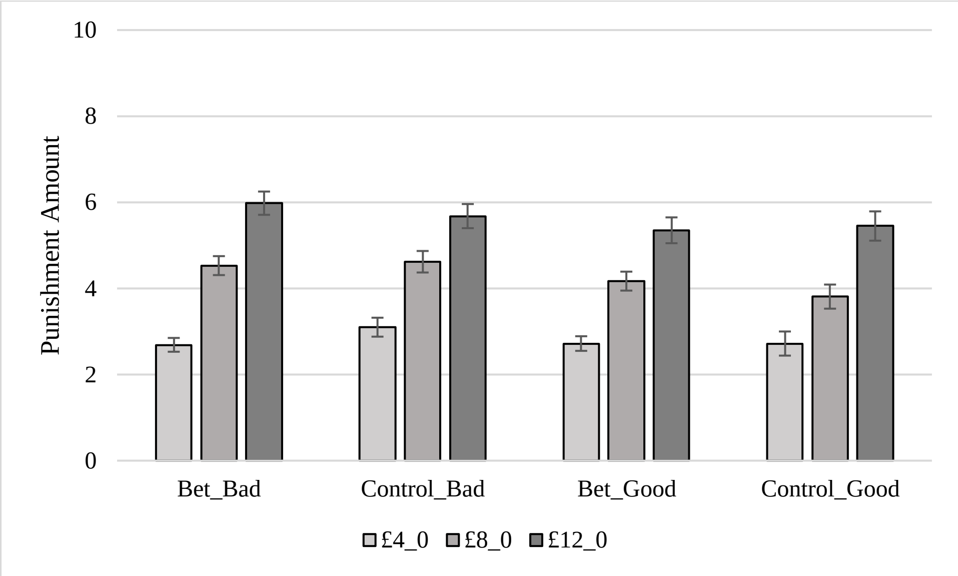
We also elicited Player A's beliefs about whether Player B would punish and if so, by how much, in each scenario. Player A received £1 if they correctly guessed the punishment amount in a randomly chosen scenario.

Figure C1: Piggyback Profit Punishment Frequency (£12-£0)



Note: # observations: Bet_Bad=251; Control_Bad=252; Bet_Good=251; Control_Good=252

Figure C2: Piggyback Profit Punishment Amount (£12-£0)



Note: # observations: Bet_Bad: £12=141; £8=139, £4=129; Control_Bad: £12=137, £8=139, £4=108; Bet_Good: £12=120, £8=111, £4=108; Control_Good: £12=84, £8=86, £4=64

Table C2: Player A's Beliefs About Piggyback Profiting Punishment (£12-£0)

VARIABLES	Bad outcome		Good outcome	
	(1)	(2)	(3)	(4)
Probability of Punishing:				
Bet	0.0458	0.0728	0.376***	0.333**
	(0.112)	(0.122)	(0.114)	(0.124)
Constant	0.110	-0.0609	-0.431***	-0.108
	(0.0791)	(0.312)	(0.0817)	(0.313)
Punishment Amount:				
Bet	0.404	0.169	-0.149	-0.761
	(0.521)	(0.564)	(0.658)	(0.660)
Constant	5.008***	5.389***	4.621***	3.436
	(0.414)	(1.534)	(0.567)	(1.766)
Sigma	1.320***	1.310***	1.342***	1.271***
	(0.0630)	(0.0664)	(0.0799)	(0.0795)
Controls	N	Y	N	Y
Observations	503	454	503	454

Notes: Cragg hurdle regressions for £12-£0 Profit. Regressions 2 and 4 include Controls. Controls include age, gender, education, political orientation, religiosity, and income. Standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05

Table C3: Player A's Beliefs About Piggyback Profiting Punishment (£8-£0 and £4-£0)

VARIABLES	Bad outcome £8		Good outcome £8		Bad outcome £4		Good outcome £4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Probability of Punishment								
Bet	0.0954	0.128	0.264*	0.333**	0.215	0.234	0.486***	0.487***
	(0.112)	(0.122)	(0.114)	(0.124)	(0.112)	(0.123)	(0.117)	(0.128)
Constant	0.0398	-0.103	-0.409***	-0.108	-0.180*	-0.371	-0.662***	-0.419
	(0.0790)	(0.311)	(0.0814)	(0.313)	(0.0794)	(0.314)	(0.0856)	(0.327)
Punishment Amount								
Bet	-0.123	-0.330	0.550	-0.761	-0.891	-1.369**	0.00796	-0.627
	(0.491)	(0.519)	(0.539)	(0.660)	(0.567)	(0.527)	(0.706)	(0.659)
Constant	3.858***	3.782**	2.922***	3.436	1.765**	3.116*	0.907	0.464
	(0.408)	(1.409)	(0.493)	(1.766)	(0.596)	(1.343)	(0.895)	(1.803)
Sigma	1.194***	1.164***	1.099***	1.271***	1.077***	0.957***	1.083***	0.963***
	(0.0706)	(0.0730)	(0.0850)	(0.0795)	(0.104)	(0.0957)	(0.131)	(0.119)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	503	454	503	454	503	454	503	454

Notes: Cragg Hurdle Model punishment for £8-£0 profit & £4-£0 profit. Regressions 2,4,6, and 8 include controls. Controls include age, gender, education, political orientation, religiosity, and income. Standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05

Table C4: Piggyback Profit Punishment Frequency. Player A (Affected Third Party)

VARIABLES	Bet Bad Outcome		Control Bad Outcome		Bet Good Outcome		Control Good Outcome	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
£12	0.0447*	0.00820	0.0922***	0.0162	0.0572**	0.0124	0.0762***	0.00410
	(0.0181)	(0.0116)	(0.0186)	(0.0128)	(0.0202)	(0.0161)	(0.0173)	(0.0136)
£8	0.0366*		0.0760***		0.0445*		0.0718***	
	(0.0181)		(0.0186)		(0.0201)		(0.0173)	
Constant	0.408***	0.446***	0.368***	0.445***	0.349***	0.395***	0.213***	0.286***
	(0.0128)	(0.00819)	(0.0131)	(0.00901)	(0.0142)	(0.0113)	(0.0122)	(0.00957)
Observations	742	493	750	499	740	491	748	496
R-squared	0.014	0.002	0.053	0.006	0.018	0.002	0.047	0.000
Number of id	251	249	252	252	251	250	252	252

Notes: LPM model with fixed effects. The dependent variable is punishment frequency. For Regressions 2,4,6,8, £8 is the Constant. For Regressions1,3,5,7, £4 is the Constant. Standard errors in parentheses*** p<0.001, ** p<0.01, * p<0.05

Table C5: Piggyback Profit Punishment Amount. Player A (Affected Third Party)

VARIABLES	Bet Bad Outcome		Control Bad Outcome		Bet Good Outcome		Control Good Outcome	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
£12	2.804***	1.491***	2.537***	1.127***	2.550***	1.434***	2.386***	1.296***
	(0.197)	(0.153)	(0.217)	(0.178)	(0.237)	(0.211)	(0.259)	(0.209)
£8	1.390***		1.410***		1.046***		1.074***	
	(0.198)		(0.215)		(0.230)		(0.258)	
Constant	1.680***	3.056***	1.962***	3.371***	2.122***	3.237***	1.457***	2.589***
	(0.141)	(0.108)	(0.156)	(0.125)	(0.165)	(0.145)	(0.186)	(0.145)
Observations	531	359	477	331	398	271	313	217
R-squared	0.375	0.355	0.315	0.204	0.328	0.276	0.312	0.283
Number of id	193	186	176	173	155	149	121	119

Notes: OLS model with fixed effects. The dependent variable is mean punishment amount, conditional on punishment occurring For Regressions 2,4,6,8, £8 is the Constant. For Regressions1,3,5,7, £4 is the Constant. Standard errors in parentheses*** p<0.001, ** p<0.01, * p<0.05

APPENDIX D: Other Regressions

Table D1: Piggyback Profit Punishment (Affected Third Party Player B)

VARIABLES	Bad outcome £8		Good outcome £8		Bad outcome £4		Good outcome £4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Probability of Punishment								
Bet	0.182	0.253	-0.0873	0.240	0.197	0.253	-0.0642	-0.0363
	(0.123)	(0.138)	(0.130)	(0.145)	(0.130)	(0.147)	(0.136)	(0.158)
Constant	-0.792***	-0.858*	-0.862***	-0.754*	-1.000***	-1.157**	-0.983***	-0.843*
	(0.0886)	(0.364)	(0.0906)	(0.375)	(0.0951)	(0.384)	(0.0945)	(0.417)
Punishment Amount								
Bet	-0.594	-0.876	-1.057	-2.135*	-0.694	-0.427	-0.711	-0.574
	(0.793)	(0.829)	(0.893)	(1.088)	(0.959)	(0.964)	(0.548)	(0.522)
Constant	2.518**	6.521**	2.810***	12.21***	0.618	4.729*	2.362***	4.888***
	(0.773)	(2.019)	(0.765)	(2.659)	(1.270)	(2.247)	(0.414)	(1.255)
Sigma	1.159***	1.043***	1.133***	1.198***	1.030***	0.899***	0.644***	0.463***
	(0.129)	(0.125)	(0.144)	(0.129)	(0.199)	(0.178)	(0.136)	(0.124)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	503	455	503	455	503	455	503	455

Hurdle Model punishment for £8-£0 profit & £4-£0 profit. Regressions 2,4,6, and 8 include controls. Controls include age, gender, education, political orientation, religiosity, and income. Standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05

Table D2: Piggyback Profit Punishment Frequency (Affected Third Party Player B)

VARIABLES	Bet Bad Outcome		Control Bad Outcome		Bet Good Outcome		Control Good Outcome	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
£12	0.0876***	0.0279	0.0556***	0	0.0797***	0.0558***	0.00794	-0.0238
	(0.0177)	(0.0164)	(0.0150)	(0.0112)	(0.0178)	(0.0166)	(0.0171)	(0.0158)
£8	0.0598***		0.0556***		0.0239		0.0317	
	(0.0177)		(0.0150)		(0.0178)		(0.0171)	
Constant								
	0.211***	0.271***	0.159***	0.214***	0.147***	0.171***	0.163***	0.194***
	(0.0125)	(0.0116)	(0.0106)	(0.00795)	(0.0126)	(0.0117)	(0.0121)	(0.0112)
Observations								
R-squared	753	502	756	504	753	502	756	504
Number of id	0.049	0.011	0.035		0.041	0.043	0.007	0.009

Notes: LPM model with fixed effects. The dependent variable is punishment frequency. Independent variables are £12-£0, £8-£0, and £4-0, respectively. For Regressions 2,4,6,8, £8 is the Constant. For Regressions1,3,5,7, £4 is the Constant. Standard errors in parentheses*** p<0.001, ** p<0.01, * p<0.05

Table D3: Piggyback Profit Punishment Amount. (Affected Third Party Player B)

VARIABLES	Bet Bad Outcome		Control Bad Outcome		Bet Good Outcome		Control Good Outcome	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
£12_£0	2.709***	1.190***	2.709***	1.080**	2.746***	1.317***	2.543***	1.026***
	(0.255)	(0.194)	(0.337)	(0.321)	(0.305)	(0.225)	(0.341)	(0.271)
£8_£0	1.544***		1.595***		1.496***		1.535***	
	(0.255)		(0.327)		(0.314)		(0.331)	
Constant	1.760***	3.124***	2.153***	3.645***	1.643***	2.859***	2.485***	3.890***
	(0.190)	(0.137)	(0.249)	(0.222)	(0.230)	(0.164)	(0.239)	(0.176)
Observations	196	143	148	108	137	100	133	92
R-squared	0.501	0.377	0.423	0.188	0.523	0.462	0.431	0.280
Number of id	81	80	58	58	61	59	57	54

Notes: LPM model with fixed effects. The dependent variable is conditional punishment amount. Independent variables are £12-£0, £8-£0, and £4-0, respectively. For Regressions 2,4,6,8, £8 is the Constant. For Regressions1,3,5,7, £4 is the Constant. Standard errors in parentheses*** p<0.001, ** p<0.01, * p<0.05

Table D4: Piggyback Profit Punishment (Unaffected Third Party Player C)

VARIABLES	Bad outcome £8		Good outcome £8		Bad outcome £4		Good outcome £4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Probability of Punishment								
Bet	-0.00719	0.0157	-0.0550	0.0595	-0.0179	-0.0275	-0.0302	-0.0519
	(0.118)	(0.123)	(0.121)	(0.124)	(0.120)	(0.125)	(0.126)	(0.134)
Constant	-0.609***	-0.0871	-0.693***	-0.122	-0.681***	-0.669	-0.850***	-0.875*
	(0.0835)	(0.352)	(0.0852)	(0.346)	(0.0850)	(0.347)	(0.0892)	(0.361)
Punishment Amount								
Bet	1.527	1.195	0.856	0.0104	-0.620	-0.471	-0.226	0.276
	(0.882)	(0.728)	(0.964)	(1.032)	(1.267)	(0.960)	(0.796)	(0.676)
Constant	0.547	0.973	0.760	1.680	-0.877	5.067	0.675	4.723*
	(1.119)	(2.150)	(1.234)	(3.074)	(2.195)	(2.664)	(1.016)	(1.837)
Sigma	1.216***	1.052***	1.255***	1.387***	1.361***	1.149***	0.943***	0.786***
	(0.137)	(0.115)	(0.152)	(0.133)	(0.207)	(0.157)	(0.179)	(0.149)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	522	522	522	522	522	522	522	522

Notes: Hurdle Model punishment for £8-£0 profit & £4-£0 profit. Regressions 2,4,6, and 8 include controls. Controls include age, gender, education, political orientation, religiosity, and income. Standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05

Table D5: Piggyback Profit Punishment Frequency (Unaffected Third Party Player C)

VARIABLES	Bet Bad Outcome		Control Bad Outcome		Bet Good Outcome		Control Good Outcome	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
£12	0.0871***	0.0606***	0.0426**	0.0194	0.0795***	0.0417*	0.0698***	0.0233
	(0.0182)	(0.0174)	(0.0154)	(0.0139)	(0.0188)	(0.0180)	(0.0179)	(0.0145)
£8	0.0265		0.0233		0.0379*		0.0465**	
	(0.0182)		(0.0154)		(0.0188)		(0.0179)	
Constant	0.242***	0.269***	0.248***	0.271***	0.189***	0.227***	0.198***	0.244***
	(0.0129)	(0.0123)	(0.0109)	(0.00986)	(0.0133)	(0.0127)	(0.0127)	(0.0102)
Observations	792	528	774	516	792	528	774	516
R-squared	0.044	0.044	0.015	0.007	0.033	0.020	0.030	0.010
Number of id	264	264	258	258	264	264	258	258

Notes: LPM model with fixed effects. The dependent variable is punishment frequency. Independent variables are £12-£0, £8-£0, and £4-0, respectively. For Regressions 2,4,6,8, £8 is the Constant. For Regressions1,3,5,7, £4 is the Constant. Standard errors in parentheses*** p<0.001, ** p<0.01, * p<0.05

Table D6: Piggyback Profit Punishment Amount (Unaffected Third Party Player C)

VARIABLES	Bet Bad Outcome		Control Bad Outcome		Bet Good Outcome		Control Good Outcome	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
£12	2.420***	1.404***	1.450***	1.212***	2.289***	1.046***	2.284***	1.017***
	(0.283)	(0.222)	(0.276)	(0.249)	(0.313)	(0.245)	(0.280)	(0.234)
£8	1.050***		0.238		1.358***		1.299***	
	(0.286)		(0.281)					
Constant	2.335***	3.249***	2.615***	2.766***	1.964***	3.162***	1.888***	3.006***
	(0.208)	(0.160)	(0.201)	(0.176)	(0.230)	(0.173)	(0.209)	(0.165)
Observations	222	158	209	145	181	131	183	132
R-squared	0.368	0.373	0.205	0.266	0.349	0.257	0.387	0.246
Number of id	92	90	79	79	79	77	75	73

Notes: OLS model with fixed effects. The dependent variable is conditional punishment amount. Independent variables are £12-£0, £8-£0, and £4-0, respectively. For Regressions 2,4,6,8, £8 is the Constant. For Regressions1,3,5,7, £4 is the Constant. Standard errors in parentheses*** p<0.001, ** p<0.01, * p<0.05

Table D7: Piggyback Profit Punishment (Moral Context Player C)

VARIABLES	Bad outcome £8		Good outcome £8		Bad outcome £4		Good outcome £4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Probability of Punishment								
Bet	-0.154	-0.141	-0.381**	-0.401**	-0.279*	-0.266*	-0.272	-0.259
	(0.124)	(0.129)	(0.137)	(0.137)	(0.126)	(0.131)	(0.147)	(0.154)
Constant	-0.630***	-0.307	-0.804***	-0.529	-0.617***	-0.422	-1.042***	-0.433
	(0.0866)	(0.353)	(0.0908)	(0.376)	(0.0864)	(0.362)	(0.0987)	(0.425)
Punishment Amount								
Bet	-1.173	-1.145	-0.816	-1.435	-3.999*	-3.602**	-0.442	-0.204
	(1.139)	(0.959)	(0.793)	(0.881)	(1.720)	(1.381)	(0.452)	(0.364)
Constant	1.455	-0.556	2.041**	7.335**	1.128	3.332	1.726***	3.439***
	(1.303)	(2.793)	(0.643)	(2.329)	(1.505)	(2.875)	(0.318)	(0.986)
Sigma	1.356***	1.206***	0.893***	1.056***	1.334***	1.204***	0.313*	0.0667
	(0.162)	(0.137)	(0.156)	(0.124)	(0.192)	(0.163)	(0.150)	(0.126)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	496	496	496	496	496	496	496	496

Notes: Hurdle Model punishment for £8-£0 profit & £4-£0 profit. Regressions 2,4,6, and 8 include controls. Controls include age, gender, education, political orientation, religiosity, and income. Standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05

Table D8: Piggyback Profit Punishment Frequency (Moral Context Player C)

VARIABLES	Bet Bad Outcome		Control Bad Outcome		Bet Good Outcome		Control Good Outcome	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
£12	0.0512***	0.0197	0.0455**	0.0496**	0.0472***	0.0236	0.0992***	0.0372*
	(0.0139)	(0.0118)	(0.0149)	(0.0152)	(0.0135)	(0.0136)	(0.0179)	(0.0159)
£8	0.0315*		-0.00413		0.0236		0.0620***	
	(0.0139)		(0.0149)		(0.0135)		(0.0179)	
Constant	0.185***	0.217***	0.269***	0.264***	0.0945***	0.118***	0.149***	0.211***
	(0.00979)	(0.00832)	(0.0105)	(0.0107)	(0.00955)	(0.00961)	(0.0127)	(0.0112)
Observations								
R-squared	762	508	726	484	762	508	726	484
Number of id	0.027	0.011	0.027	0.043	0.024	0.012	0.061	0.022

Notes: LPM model with fixed effects. The dependent variable is punishment frequency. Independent variables are £12-£0, £8-£0, and £4-0, respectively. For Regressions 2,4,6,8, £8 is the Constant. For Regressions1,3,5,7, £4 is the Constant. Standard errors in parentheses*** p<0.001, ** p<0.01, * p<0.05

Table D9: Piggyback Profit Punishment Amount. Player C (Moral Context Player C)

VARIABLES	Bet Bad Outcome		Control Bad Outcome		Bet Good Outcome		Control Good Outcome	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
£12	2.237***	0.925**	1.644***	1.651***	2.067***	0.889*	2.756***	1.208***
	(0.323)	(0.300)	(0.292)	(0.226)	(0.345)	(0.343)	(0.263)	(0.240)
£8	1.279***		0.106		1.083**		1.516***	
	(0.319)		(0.306)		(0.340)		(0.265)	
Constant	1.947***	3.144***	3.406***	3.418***	1.379***	2.394***	1.492***	2.860***
	(0.236)	(0.213)	(0.213)	(0.163)	(0.256)	(0.243)	(0.203)	(0.171)
Observations	162	115	205	140	90	66	147	111
R-squared	0.331	0.155	0.243	0.462	0.423	0.205	0.574	0.350
Number of id	63	62	77	77	39	39	63	63

Notes: PLAYER C: LPM model with fixed effects. The dependent variable is conditional punishment amount. Independent variables are £12-£0, £8-£0, and £4-0, respectively. For Regressions 2,4,6,8, £8 is the Constant. For Regressions1,3,5,7, £4 is the Constant. Standard errors in parentheses*** p<0.001, ** p<0.01, * p<0.05

Table D10: Mean Equivalence Tests for Punishment Frequency (Study 2: Unaffected Third Party)

	H0: The Bet treatment decreases/increases punishment frequency by x percentage points					
x	-5	5	-7.5	7.5	-10	10
£8 Bet v Control (Bad Outcome)	1.23	1.35*	1.87**	1.99**	2.51***	2.63***
£4 Bet v Control (Bad Outcome)	1.12	1.48*	1.84**	2.14**	2.51***	2.81***
£8 Bet v Control (Good Outcome)	0.89	1.80**	1.56*	2.45***	2.24**	3.15***
£4 Bet v Control (Good Outcome)	1.21	1.67**	1.93**	2.41***	2.65***	3.13***

Note: Z-scores reported above. All t-statistics for TOST procedures on a variety of lower and upper equivalence bounds (in standardized coefficients). *** p<0.01, ** p<0.05, * p<0.1.

Table D11: Mean Equivalence Tests for Punishment Amount (Study 2: Unaffected Third Party)

	H0: The Bet treatment decreases/increases punishment amount by x points					
x	-0.5	0.5	-1	1	-1.5	1.5
£12 Bet v Control (Bad Outcome)	2.17**	-0.03	3.24***	1.04	4.30***	2.12**
£8 Bet v Control (Bad Outcome)	3.10***	-0.51	4.38***	0.77	5.67***	2.07**
£4 Bet v Control (Bad Outcome)	0.82	1.80**	2.13**	3.11***	3.44***	4.42***
£12 Bet v Control (Good Outcome)	1.03	1.02	2.05**	2.04**	3.08***	3.07***
£8 Bet v Control (Good Outcome)	2.07**	0.29	3.25***	1.50*	4.42***	2.64***
£4 Bet v Control (Good Outcome)	1.22	1.78**	2.73***	3.30***	4.24***	4.80***

Notet: t-stat reported above. All t-statistics for TOST procedures on a variety of lower and upper equivalence bounds (in standardized coefficients). *** p<0.01, ** p<0.05, * p<0.1.

APPENDIX E: Study 2 Player A Results

E1: Player A's betting decisions

Table E1 reports the betting decisions of Player A in *Bet*. Consistent with the results from affected third parties, Player A is more likely to bet that Player B has a good outcome than a bad outcome in the RPS game for £12 (binomial test, p=0.025),[74] £8 (binomial test, p=0.021), and £4 (p=0.005). We also find that Player A is more likely to bet on a good outcome than a bad outcome when the profit is £0 (p=0.042). Once again, betting is also much more likely when there is a positive profit compared to no profit (£4 v £0: 66.7% vs 29.9%, McNemar's test, p<0.001).[75] However, we do observe about 30% of participants who bet even when there is no profit. These results are consistent with Player A's behaviour in Study 1.

We also asked Players A if they thought Player B would be more likely to lose, win, or equally likely to lose or win the game of RPS. The majority of participants (70.0%) stated that Player B was just as likely to lose as win. For the remaining participants, the proportion of those who thought Player B was more likely to lose is slightly higher than that of those who thought Player B was more likely to win, but the difference is not statistically significant (17.4% vs.12.5%, binomial test, p=0.176). This is consistent with the findings of beliefs in Study 1.

Table E1: Betting decisions of Player A in *Bet* (Third Party)

Profit	**Bet on bad outcome**	**Bet on good outcome**	**No bet**
12	34.9%	47.7%	17.4%
8	34.5%	47.7%	17.8%
4	26.1%	40.5%	33.3%
0	11.4%	18.6%	70.1%

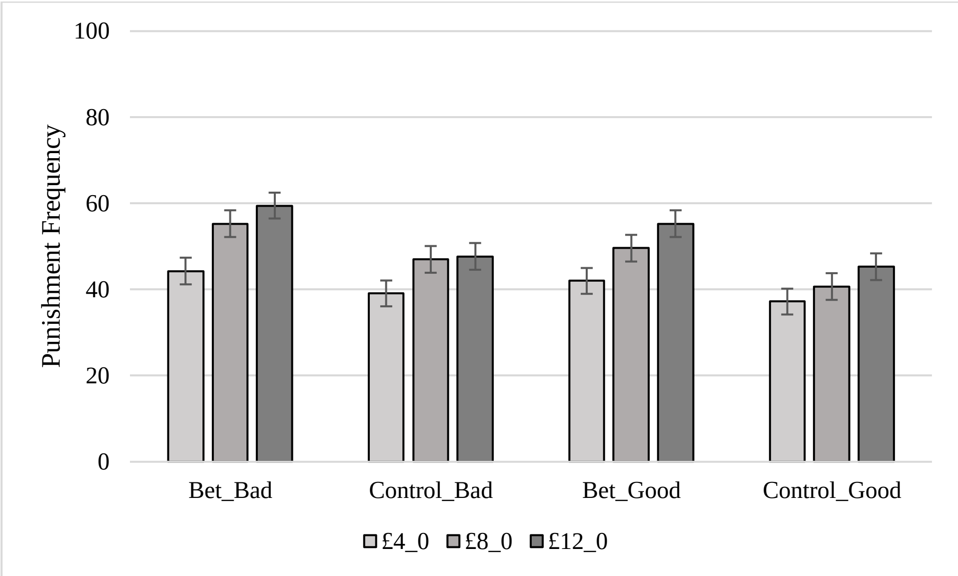
E2: Player A's beliefs about punishment

[74] The null hypothesis for the binomial test is that the probability of betting on either winning or losing is p=0.5

[75] We also find p<0.001 for £12 v £0, and £8 v £0.

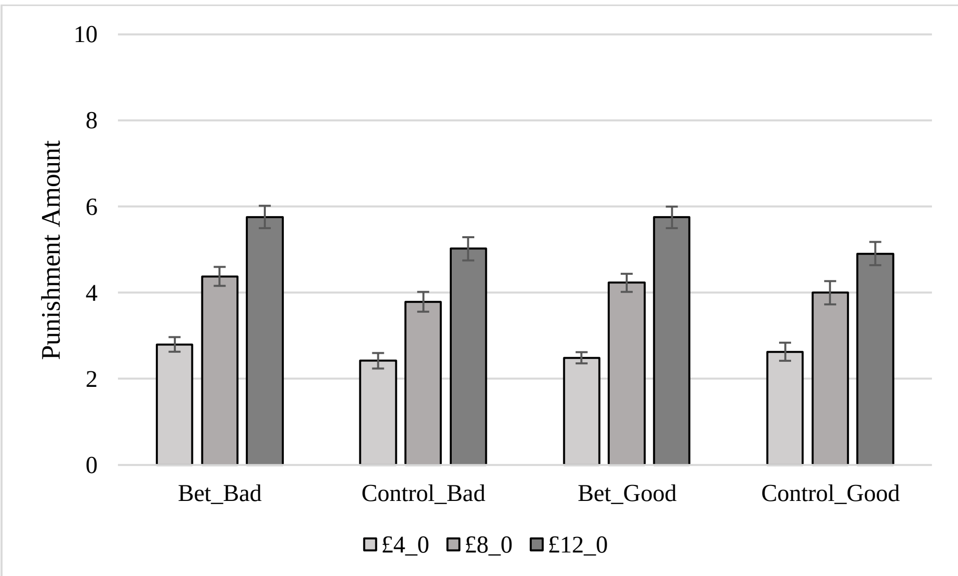
We again elicited Player A's beliefs about whether Player C will punish them in each scenario in an incentive compatible manner.

Figure E1: Piggyback Profit Punishment Frequency



Note: # observations: Bet_Bad=264; Control_Bad=258; Bet_Good=264; Control_Good=258

Figure E2: Piggyback Profit Punishment Amount



Note: # observations: Bet_Bad: £12=157; £8=146, £4=117; Control_Bad: £12=123, £8=121, £4=101; Bet_Good: £12=146, £8=131, £4=111; Control_Good: £12=117, £8=105, £4=96

TABLE E2: Player A's Beliefs About Piggyback Profiting Punishment £12-£0 (Unaffected Third Party)

VARIABLES	Bad outcome		Good outcome	
	(1)	(2)	(3)	(4)
Probability of Punishing:				
Bet	0.298**	0.302**	0.250*	0.281*
	(0.110)	(0.115)	(0.110)	(0.115)
Constant	-0.0583	0.787*	-0.117	0.527
	(0.0781)	(0.315)	(0.0782)	(0.312)
Punishment Amount:				
Bet	1.050*	1.071*	1.157*	1.228*
	(0.534)	(0.543)	(0.518)	(0.521)
Constant	4.117***	6.399***	4.101***	5.826***
	(0.466)	(1.497)	(0.444)	(1.434)
Sigma	1.304***	1.266***	1.256***	1.217***
	(0.0656)	(0.0636)	(0.0657)	(0.0638)
Controls	N	Y	N	Y
Observations	522	521	522	521

Notes: Cragg hurdle regressions for £12-£0 Profit. Regressions 2 and 4 include Controls. Controls include age, gender, education, political orientation, religiosity, and income. Standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05

TABLE E3: Player A's Beliefs About Piggyback Profiting Punishment £8-£0 and £4-0 (Unaffected Third Party)

VARIABLES	Bad outcome £8		Good outcome £8		Bad outcome £4		Good outcome £4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Probability of Punishment								
Bet	0.211	0.202	0.226*	0.281*	0.133	0.123	0.126	0.131
	(0.110)	(0.115)	(0.110)	(0.115)	(0.111)	(0.116)	(0.111)	(0.116)
Constant	-0.0778	0.560	-0.235**	0.527	-0.275***	0.638*	-0.326***	0.384
	(0.0781)	(0.312)	(0.0788)	(0.312)	(0.0791)	(0.320)	(0.0796)	(0.318)
Punishment Amount								
Bet	0.962	0.969	0.374	1.228*	0.798	0.797	-0.252	-0.0983
	(0.519)	(0.499)	(0.551)	(0.521)	(0.515)	(0.452)	(0.427)	(0.360)
Constant	2.592***	3.478*	2.911***	5.826***	0.884	2.457*	1.773***	1.142
	(0.501)	(1.412)	(0.520)	(1.434)	(0.613)	(1.227)	(0.416)	(1.035)
Sigma	1.180***	1.119***	1.189***	1.217***	0.948***	0.834***	0.819***	0.666***
	(0.0768)	(0.0726)	(0.0823)	(0.0638)	(0.105)	(0.0927)	(0.0975)	(0.0841)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	522	521	522	521	522	521	522	521

Notes: Cragg Hurdle Model punishment for £8-£0 profit & £4-£0 profit. Regressions 2,4,6, and 8 include controls. Controls include age, gender, education, political orientation, religiosity, and income. Standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05

Table E4: Player A's Beliefs About Piggyback Punishment Frequency. (Unaffected Third Party)

VARIABLES	Bet Bad Outcome		Control Bad Outcome		Bet Good Outcome		Control Good Outcome	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
£12	0.121***	0.0311	0.101***	0.0245	0.147***	0.0508***	0.0758***	0.0403*
	(0.0208)	(0.0164)	(0.0196)	(0.0152)	(0.0189)	(0.0148)	(0.0185)	(0.0169)
£8	0.0889***		0.0762***		0.0965***		0.0358	
	(0.0207)		(0.0195)		(0.0189)		(0.0185)	
Constant	0.362***	0.451***	0.349***	0.427***	0.311***	0.408***	0.334***	0.370***
	(0.0147)	(0.0116)	(0.0137)	(0.0107)	(0.0133)	(0.0105)	(0.0130)	(0.0120)
Observations	779	519	747	496	780	519	755	502
R-squared	0.066	0.014	0.056	0.010	0.108	0.044	0.033	0.022
Number of id	263	262	252	251	264	263	255	254

Notes: LPM model with fixed effects. The dependent variable is punishment frequency. Independent variables are £12-£0, £8-£0, and £4-0, respectively. For Regressions 2,4,6,8, £8 is the Constant. For Regressions 1,3,5,7, £4 is the Constant. Standard errors in parentheses*** p<0.001, ** p<0.01, * p<0.05

Table E5: Player A's Beliefs About Piggyback Profit Punishment Amount. (Unaffected Third Party)

VARIABLES	Bet Bad Outcome		Control Bad Outcome		Bet Good Outcome		Control Good Outcome	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
£12	3.222***	1.407***	2.472***	1.081***	3.201***	1.632***	2.405***	1.280***
	(0.208)	(0.180)	(0.210)	(0.190)	(0.208)	(0.179)	(0.213)	(0.188)
£8	1.788***		1.344***		1.556***		1.149***	
	(0.206)		(0.207)		(0.207)		(0.213)	
Constant	1.777***	3.486***	1.584***	2.904***	1.609***	3.144***	1.809***	2.886***
	(0.151)	(0.126)	(0.151)	(0.132)	(0.151)	(0.127)	(0.153)	(0.132)
Observations	509	357	419	292	468	329	378	260
R-squared	0.432	0.269	0.349	0.194	0.457	0.356	0.356	0.285
Number of id	193	190	159	156	182	177	145	142

Notes: OLS model with fixed effects. The dependent variable is conditional punishment amount. Independent variables are £12-£0, £8-£0, and £4-0, respectively. For Regressions 2,4,6,8, £8 is the Constant. For Regressions1,3,5,7, £4 is the Constant. Standard errors in parentheses*** p<0.001, ** p<0.01, * p<0.05

APPENDIX F: Study 3 Player A Results

Table F1 below reports the betting decisions of Player in *Bet*. We find that in all positive profit scenarios, Player A is more likely to bet on more than zero deaths than on zero deaths. This is the case in the £12 profit condition (binomial test, p<0.001), £8 (binomial test, p<0.001), £4 (binomial test, p=0.001), and £0 (binomial test, p=0.003).[76] Betting is also much more likely to happen when there is a positive profit compared to zero profit (£4 v £0: 74.5% vs 29.9%, McNemar's test, p<0.001).[77]

We find that the majority of Player As believe it is more likely that at least one person will die (59.1%). Conditional on Player A believing one of the outcomes was more likely, significantly more participants believed that at least one person would die is more likely than zero deaths (77.3% vs 22.7%, Binomial test, p<0.001). When the profit is £12, 89.3% of participants who thought it was more likely there would be at least one death bet on this option. 72.7% of participants who thought zero deaths was more likely bet on zero deaths. Finally, of those who said the two outcomes were equally likely, 46.7% of them bet on more than deaths and 15.0% bet on zero deaths. These results suggest that if a bet was made, the betting decision was mostly driven by the participants' belief about what was more likely to occur.

Table F1: Player A's Betting Decisions (Moral context)

Profit	**Bet on >0 deaths**	**Bet on 0 deaths**	**No bet**
£12	65.4%	18.5%	16.1%
£8	62.2%	20.1%	17.7%
£4	52.0%	18.1%	29.9%
£0	16.9%	7.5%	75.6%

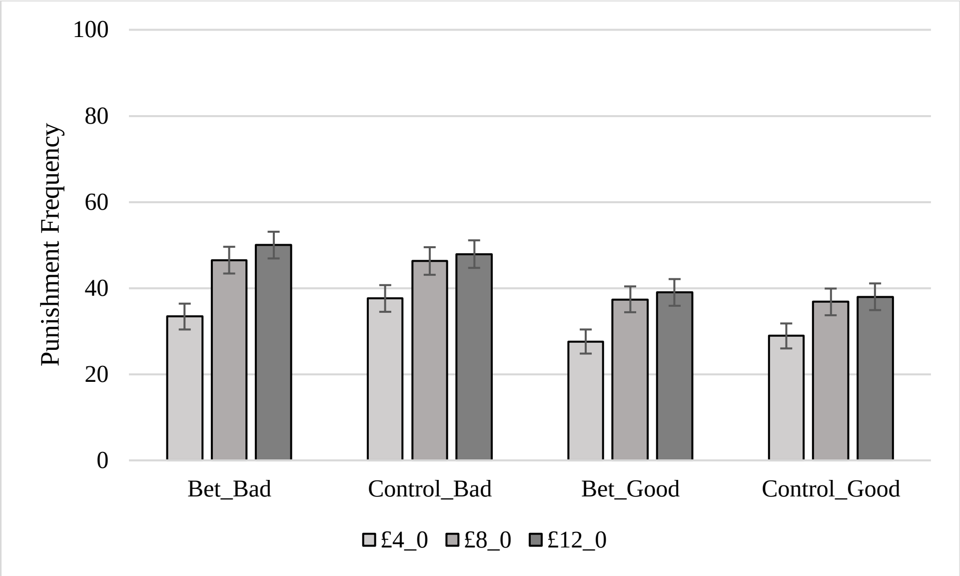
F2: Player A's beliefs about punishment

We again elicited Player A's beliefs about whether Player C will punish them in each scenario in an incentive compatible manner.

[76] The binomial test is compared to a random decision of 0.5 likelihood of each outcome.

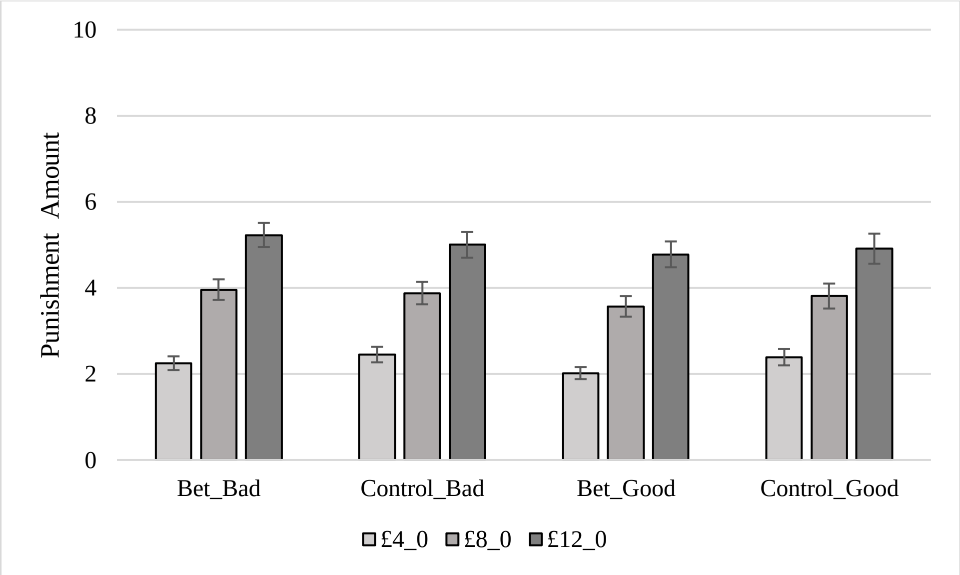
[77] We also find p<0.001 for £12 v £0, and £8 v £0.

Figure F1: Player A's Beliefs About Piggyback Profit Punishment Frequency



Note: # observations: Bet_Bad=254; Control_Bad=242; Bet_Good=254; Control_Good=242

Figure F2: Player A's Beliefs About Piggyback Profit Punishment Amount



Note: # observations: Bet_Bad: £12=127; £8=118, £4=85; Control_Bad: £12=116, £8=112, £4=91; Bet_Good: £12=88, £8=95, £4=70; Control_Good: £12=92, £8=89, £4=70

TABLE F2: Player A's Beliefs About Piggyback Profiting Punishment £12-£0 (Moral Context)

VARIABLES	Bad outcome		Good outcome	
	(1)	(2)	(3)	(4)
Probability of Punishment:				
Bet	0.0518	0.0546	0.0251	0.0317
	(0.113)	(0.117)	(0.114)	(0.118)
Constant	-0.0518	0.322	-0.305***	-0.370
	(0.0806)	(0.298)	(0.0819)	(0.303)
Punishment Amount:				
Bet	0.372	0.277	-0.239	-0.254
	(0.669)	(0.651)	(0.808)	(0.790)
Constant	3.673***	3.475*	3.387***	3.353
	(0.617)	(1.670)	(0.772)	(2.069)
Sigma	1.405***	1.359***	1.433***	1.373***
	(0.0811)	(0.0777)	(0.0989)	(0.0931)
Controls	N	Y	N	Y
Observations	496	496	496	496

Notes: Cragg hurdle regressions for £12-£0 Profit. Regressions 2 and 4 include Controls. Controls include age, gender, education, political orientation, religiosity, and income. Standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05

TABLE F3: Player A's Beliefs About Piggyback Profiting Punishment (Moral context)

VARIABLES	Bad outcome £8		Good outcome £8		Bad outcome £4		Good outcome £4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Probability of Punishment								
Bet	0.00442	0.0324	0.0165	0.0317	-0.111	-0.105	-0.0404	-0.0165
	(0.113)	(0.117)	(0.115)	(0.118)	(0.116)	(0.119)	(0.120)	(0.125)
Constant	-0.0934	0.0892	-0.338***	-0.370	-0.316***	-0.122	-0.556***	-0.250
	(0.0807)	(0.298)	(0.0823)	(0.303)	(0.0820)	(0.305)	(0.0853)	(0.321)
Punishment Amount								
Bet	0.164	-0.0256	-0.492	-0.254	-0.418	-0.457	-0.586	-0.917**
	(0.697)	(0.636)	(0.769)	(0.790)	(0.503)	(0.458)	(0.389)	(0.344)
Constant	2.054**	2.562	2.105*	3.353	1.391**	1.117	1.938***	2.313*
	(0.758)	(1.684)	(0.821)	(2.069)	(0.520)	(1.228)	(0.317)	(0.900)
Sigma	1.323***	1.237***	1.290***	1.373***	0.839***	0.733***	0.570***	0.399***
	(0.0995)	(0.0909)	(0.115)	(0.0931)	(0.117)	(0.104)	(0.107)	(0.0927)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	496	496	496	496	496	496	496	496

Notes: Hurdle Model punishment for £8-£0 profit & £4-£0 profit. Regressions 2,4,6, and 8 include controls. Controls include age, gender, education, political orientation, religiosity, and income. Standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05

Table F4: Player A's Beliefs About Piggyback Profit Punishment Frequency, Moral Context

VARIABLES	Bet Bad Outcome		Control Bad Outcome		Bet Good Outcome		Control Good Outcome	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
£12	0.138***	0.0277*	0.0828***	0.0124	0.103***	0.0119	0.0671***	0.00844
	(0.0194)	(0.0118)	(0.0184)	(0.0138)	(0.0190)	(0.0143)	(0.0165)	(0.0146)
£8	0.110***		0.0702***		0.0906***		0.0583***	
	(0.0193)		(0.0184)		(0.0190)		(0.0164)	
Constant	0.296***	0.406***	0.352***	0.422***	0.245***	0.336***	0.267***	0.326***
	(0.0137)	(0.00830)	(0.0130)	(0.00972)	(0.0134)	(0.0101)	(0.0116)	(0.0103)
Observations	761	507	725	483	760	506	719	478
R-squared	0.101	0.022	0.047	0.003	0.065	0.003	0.040	0.001
Number of id	254	254	242	242	254	254	242	241

Notes: LPM model with fixed effects. The dependent variable is punishment frequency. Independent variables are £12-£0, £8-£0, and £4-0, respectively. For Regressions 2,4,6,8, £8 is the Constant. For Regressions 1,3,5,7, £4 is the Constant. Standard errors in parentheses*** p<0.001, ** p<0.01, * p<0.05

Table F5: Player A's Beliefs About Piggyback Profit Punishment Amount, (Moral Context)

VARIABLES	Bet Bad Outcome		Control Bad Outcome		Bet Good Outcome		Control Good Outcome	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
£12	2.876***	1.331***	2.800***	1.298***	2.709***	1.287***	2.914***	1.313***
	(0.178)	(0.135)	(0.196)	(0.148)	(0.195)	(0.148)	(0.211)	(0.176)
£8	1.558***		1.493***		1.428***		1.590***	
	(0.178)		(0.194)		(0.193)		(0.206)	
Constant	1.427***	2.995***	1.639***	3.061***	1.286***	2.722***	1.256***	2.826***
	(0.132)	(0.0958)	(0.142)	(0.104)	(0.144)	(0.103)	(0.150)	(0.122)
Observations	419	300	372	260	325	232	311	215
R-squared	0.500	0.402	0.468	0.384	0.493	0.413	0.500	0.362
Number of id	156	155	138	136	125	124	118	116

Notes: OLS model with fixed effects. The dependent variable is punishment amount. Independent variables are £12-£0, £8-£0, and £4-0, respectively. For Regressions 2,4,6,8, £8 is the Constant. For Regressions 1,3,5,7, £4 is the Constant. Standard errors in parentheses*** p<0.001, ** p<0.01, * p<0.05

APPENDIX G: Negative Beliefs

Player A holds a negative belief if they think there is $p > 0.5$ that Player B loses the RPS game. If negative beliefs are punished, then we should expect B (C) to punish A when A bets that B will have a bad outcome in the RPS game even though there is no profit from the bet. Therefore, B (C) would be more likely to punish A in this case than when A bets (for zero profit) that B has a good outcome.

Thus, by comparing $F(\pi > 0, g)$ and $F(\pi = 0, g)$, we can test to what extent the punishment may be driven by the beliefs and whether it creates confounds for the testing of the hypotheses. We report these comparisons for punishment frequency (Table G1) and average punishment amount below (Table G2). Across all comparisons, we fail to find any evidence that suggests negative beliefs affect punishment behaviour.

Table G1: Negative Beliefs Punishment Frequency

	£0 Bet Bad	£0 Bet Good	p value
Affected Third Party	12.0%	10.8%	p=0.629
Unaffected Third Party	14.0%	11.7%	p=0.210
Moral	11.8%	8.7%	p=0.115

Note: p-values are calculated using McNemar's test

Table G2: Negative Beliefs Average Amount of Punishment

	£0 Bet Bad	£0 Bet Good	p value
Affected Third Party	4.85	4.50	p=0.094
Unaffected Third Party	3.16	2.76	p=0.296
Moral	3.23	2.45	p=0.155

Note: p-values are calculated using Wilcoxon's signed-rank test