

Arregui-Alegria, Iker; Gokmen, Gunes; van Veldhuizen, Roel

Working Paper

Public discourse, norm adherence, and socially responsible behavior

Working Paper, No. 2025:10

Provided in Cooperation with:

Department of Economics, School of Economics and Management, Lund University

Suggested Citation: Arregui-Alegria, Iker; Gokmen, Gunes; van Veldhuizen, Roel (2025) : Public discourse, norm adherence, and socially responsible behavior, Working Paper, No. 2025:10, Lund University, School of Economics and Management, Department of Economics, Lund

This Version is available at:

<https://hdl.handle.net/10419/339055>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



LUND UNIVERSITY

Public Discourse, Norm Adherence, and Socially Responsible Behavior

Arregui Alegria, Iker; Gokmen, Gunes; van Veldhuizen, Roel

2025

Document Version:
Other version

[Link to publication](#)

Citation for published version (APA):

Arregui Alegria, I., Gokmen, G., & van Veldhuizen, R. (2025). *Public Discourse, Norm Adherence, and Socially Responsible Behavior*. (Working Papers; No. 2025:10).

Total number of authors:
3

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Working Paper 2025:10

Department of Economics
School of Economics and Management

Public Discourse, Norm Adherence, and Socially Responsible Behavior

Iker Arregui-Alegria
Gunes Gokmen
Roel van Veldhuizen

Dec 2025



LUND
UNIVERSITY

Public Discourse, Norm Adherence, and Socially Responsible Behavior*

Iker Arregui-Alegria[†] Gunes Gokmen[‡] Roel van Veldhuizen[§]

December 12, 2025

Abstract

This paper examines the causal effect of public discourse in one or both sides of the market on overall market socially responsible behavior. In a laboratory setting, we vary whether firms and/or consumers participate in a public discussion before trading begins. When both sides take part, the share of socially responsible trades increases slightly; however, when only one side participates, market social responsibility does not improve relative to a no-discourse control. These findings suggest that campaigns aiming to foster socially responsible conduct must engage all sides of the market to achieve meaningful impact. We also provide evidence that the effectiveness of public discourse will be limited when participants prioritize profits over norm adherence.

Keywords: Public Discourse, Externalities, Social Responsibility, Experiment.

JEL classification codes: C92, D62, D83, M14.

*The study was pre-registered at the AEA RCT registry as trial 15202. We would like to thank Björn Bartling, Ernst Fehr, Kevin McCabe, Roberto Weber and seminar participants at Zurich, George Mason, Lund, Aarhus, and the Copenhagen Network for Experimental Economists for valuable comments. This research benefited from financial support from the Jan Wallander and Tom Hedelius Foundation.

[†]Lund University; email: iker.arregui.alegria@nek.lu.se.

[‡]Lund University; email: gunes.gokmen@nek.lu.se.

[§]Lund University; email: roel.van_veldhuizen@nek.lu.se.

1 Introduction

Consumers, firms, and public agencies increasingly incorporate broader social considerations into their routine market decisions. For example, shoppers pay premiums for locally sourced produce, companies replace conventional cheap plastic packaging with more sustainable and more expensive containers, and municipalities invest in low-emission public transport. Public campaigns often reinforce these shifts by fostering open discussion of the harms associated with standard market behavior. While existing evidence indicates that discourse involving all parties can promote more responsible conduct (Bartling et al., 2024b), it remains unclear whether similar changes arise when conversation is limited to only one side of a transaction. *Will prompting only consumers (or producers) to debate the environmental costs of single-use plastics induce producers (consumers) to offer (buy) greener packaging alternatives?* To address this question, the present study introduces one-sided public discourse in a large controlled laboratory market experiment (N=483) and examines whether targeted dialogue can propagate across the market, affecting trading behavior and perceived social norms.

The impact of public discourse on socially responsible behavior may depend on which side of the market participates. This matters because many real-world campaigns—ranging from sustainability initiatives to corporate social responsibility drives—often target only one side of the market, leaving it unclear whether asymmetric engagement can shift outcomes (Crane et al., 2019). In theory, consumer discourse alone could suffice: buyers hold substantial market power through their purchasing decisions; and if they demand only socially responsible products, sellers must respond (Micheletti, 2003; Smith, 2008). Yet, it is the sellers who determine the range of available products, and they often face competitive pressures that discourage responsible conduct (Baron, 2001), suggesting they might require discourse more directly. It is therefore not obvious which side’s engagement could be more effective. By comparing buyer-only, seller-only, and joint discourses, our study directly tests

whether ‘conversation’ on one or both sides of the market can translate into action in the form of more socially responsible trade.

In our study, we simulate a market setting where four sellers and three buyers can trade with one another across twenty-four rounds. Every round, each seller first chooses which of two products to offer: a low-cost “harmful” good that reduces a common donation fund—thereby imposing a monetary loss on an uninvolved charity—or a high-cost “responsible” good that leaves the fund intact. Buyers then decide whether, and from whom, to purchase. Before trading begins, we vary the scope of public discourse: no discourse, discourse confined to sellers, discourse confined to buyers, or discourse open to both sides. Throughout the session, we record prices, profits, and the share of responsible trades, and elicit participants’ pre- and post-market norms (judgments of the appropriateness of buying the harmful good, Krupka and Weber, 2013). We also classify chat content with a large-language-model framework. This design allows us to determine whether discourse restricted to a single side of the market can spill over to influence behavior and social norms, and to benchmark these effects against those produced by full, two-sided communication.

Our public-discourse manipulation consists of a single, five-minute, free-form chat conducted immediately before the first trading round; we replicate this protocol exactly but vary *who* is allowed to participate: sellers, buyers, or both. Our design builds on Bartling et al. (2024b), who compare a *Full Discourse* treatment that gives voice to all participants to a *No Discourse* treatment that removes it altogether. We extend their framework by also including a *Seller-Only* and *Buyer-Only* treatment that restrict the same conversation channel to a single side of the market. This structure allows us to separate the behavioral and normative impact of (*full*) unrestricted discourse from that of (*partial*) restricted dialogue—an aspect left unexplored in the broader experimental literature on social responsibility. By embedding these variants in a repeated market that requires traders to forgo private earnings to reduce a third-party loss, our study links the prosocial effects of communication documented in earlier work (Dawes et al., 1977; Charness and Dufwenberg, 2006; Bochet et al., 2006) to

the problem of external-cost mitigation and tests whether targeted public discourse can propagate beyond its immediate audience to reshape marketwide behavior.

Our first result, where both the sellers and buyers are treated in the discourse condition, *Full Discourse*, is in line with the pro-social effect documented by Bartling et al. (2024b). However, in our setting, the effect size is at roughly one quarter of the magnitude found by Bartling et al. (2024b): the share of responsible trades rises by about 8 percentage points relative to the no-discourse baseline ($p \approx 0.09$, one-sided), compared with the 29-point jump they observe.¹ Second, we find that when dialogue is restricted to a single side of the market, public discourse is no longer effective. Neither the *Seller-Only* nor the *Buyer-Only* treatment differs significantly from the no-discourse condition in terms of the share of responsible trades. We also do not observe a consistent effect of any of our discourse treatments on the elicited social norms.

Why did our *Full Discourse* intervention have a much smaller effect (8pp) than the one in Bartling et al. (2024b) (29pp)? First, although both set of participants view trading harmful products as somewhat inappropriate, we show that our participants find it slightly more acceptable to trade harmful products. Second, we show that our participants, unlike those in Bartling et al. (2024b), prioritize profits over norm adherence. We additionally carry out content analysis to reveal that participants discuss similar topics in both studies. Taken together, these results suggest that the smaller effect of public discourse is driven by a combination of differences in baseline social norms, and differences in the importance attached to those norms when making decisions. This highlights an important boundary condition: the effectiveness of public discourse depends on both norms of responsible behavior per se and the extent to which individuals value such norms relative to material payoffs.

In this paper, we first contribute to a growing experimental literature that studies the conditions under which individuals are willing to internalize externalities (Rode et

¹This is the effect observed in Study 2, which most closely resembles our experiment (Figure 3 on page 3056 in Bartling et al., 2024b). They observe larger effects in Study 1 with both Chinese and Swiss participants.

al., 2008; Bénabou and Tirole, 2010; Dufwenberg et al., 2011, 2022; Bartling et al., 2015, 2019, 2024a). More concretely, we contribute to this line of work by disentangling which market actors must be exposed to normative appeals that induce players to internalize externalities before such restraint creates spill-over effects in a market as a whole. We also contribute to this line of work by replicating a high-profile recent study (Bartling et al., 2024b), showing that the effectiveness of public discourse interventions may depend on the underlying population, particularly the extent to which the population values norm adherence over profit maximization. Consequently, the impact of discourse observed in one setting may not generalize to contexts where competitive pressures or profit motives dominate normative concerns. From a policy perspective, this suggests that public campaigns may be most effective in markets where social norms are already salient, but may require complementary regulation or incentives in settings where profit motives strongly outweigh normative considerations.

Second, decades of laboratory evidence link pre-play communication to higher cooperation in social dilemmas, coordination games, and incomplete contracts (Dawes et al., 1977; Charness and Dufwenberg, 2006; Bochet et al., 2006). Yet, these settings typically involve Pareto-improving agreements that benefit every speaker. Our experiment tackles the harder case in which acting prosocially is privately costly and benefits only a third party. We show that when conversation is limited to a single side of the market, the cooperative boost disappears—echoing findings from dictator games that asymmetric messaging can dampen generosity (Andreoni and Rao, 2011) and complementing evidence from bargaining contexts where communication may entrench rather than redress inequities (Agranov and Tergiman, 2014). By demonstrating that “some talk” is not always “better than none,” our study refines the conditions under which discourse can be expected to produce socially desirable outcomes in markets. This suggests that real-world campaigns—say, for sustainability or corporate social responsibility—may fall short if they target only firms or only consumers,

and that fostering dialogue across both sides of the market may be necessary to generate meaningful behavioral change.

This paper proceeds as follows. Section 2 provides information on the laboratory experiment, including the design and the main predictions. Section 3 provides the main results, and in Section 4, we examine the mechanisms. Section 5 concludes.

2 Experimental Design

Our experiment builds on Bartling et al. (2024b) and features a product market where participants are assigned roles as sellers or buyers that remain constant across 24 periods. One period is randomly selected for payment. We vary whether sellers and/or buyers can engage in public discourse prior to the first period using a 2x2 design, and elicit social norms before period 1 and after period 24. In what follows, we will explain each of these design elements in greater detail.

The Market— We consider experimental markets that consist of seven participants, four sellers and three buyers. At the beginning of the session, participants are randomly assigned to one of these roles. Each participant remains in the same market and in the same role for the duration of the experiment (24 periods).

At the beginning of each period, sellers decide between producing one of the two products. The first one is a harmful product, which costs 0 points to produce but generates a negative externality by reducing the donation fund by 600 points. The second one is a responsible product, which costs 200 points to produce and does not affect the donation fund. Both products yield the same value of 500 points to the buyer. Thus, the products differ only in terms of their production costs and their impact on the donation fund.

After sellers choose their preferred product type and set a price, buyers enter the market one at a time in random order. Each buyer observes the product type and price of all available

products. They then decide whether to purchase a product or abstain from buying. If they decide to buy a product, then that product is removed from the market and the next buyer gets to pick a product from those that are still available (see Figure D.3 in the Appendix as an illustration). During this process, sellers receive live updates whenever a buyer makes a decision, learning exactly which product was bought, its price, its type, and whether it was their own or another seller's.

After all buyers have made their decisions, earnings are calculated and participants learn their individual earnings and the total amount donated to the charity across all transactions. Seller earnings are equal to the price minus production costs, and buyers earn 500 points minus the price they paid. Sellers and buyer who do not complete a trade receive 0 points. Total earnings for the charity are reduced by 600 points for every completed trade featuring the harmful product, from a total of 3000 points per period per market. For example, if all traded products in a period are harmful, the total donation for that period will be 1,200 points (see Figure D.4 in the Appendix for an illustration). We employed the charity GiveDirectly, a charity that provides direct cash transfers to individuals living in extreme poverty.

Treatments— In our experiment, participants are randomly assigned to one of four treatment conditions using a 2×2 between-subjects design. We vary whether (a) sellers and (b) buyers are allowed to engage in public discourse before the first round of the market game. This results in four conditions: (i) No Discourse, where neither sellers nor buyers can communicate; (ii) Full Discourse, where all sellers and buyers in a given market can chat with one another; (iii) Seller-Only Discourse, where only sellers can chat among themselves; and (iv) Buyer-Only Discourse, where only buyers can chat among themselves.

In the discourse conditions, participants engage in a one-time, five-minute chat session before the first period. Importantly, when the chat takes place, participants do not yet know their role in the experiment (buyer or seller). Following Bartling et al. (2024b), the

chat interface asks participants to discuss the social appropriateness of trading the harmful product. Beyond this prompt, communication is unrestricted, provided messages do not contain personally identifying, obscene, or insulting content. In the No Discourse condition, participants proceed directly to the market game after being informed of their assigned roles.

Social Norm Elicitation— After the chat, participants individually evaluate how socially appropriate or inappropriate they consider trading the harmful product to be. They choose from the following four statements: very socially appropriate, somewhat socially appropriate, somewhat socially inappropriate, and very socially inappropriate. Following Krupka and Weber (2013), participants are then asked to predict the modal response on the previous question in their session, receiving 25 points for a correct prediction. These evaluations (first step) and predictions (second step) are repeated after period 24.

Procedures— We conducted the study at LINEEX in Valencia, Spain. The experiment took place over eight sessions across three days, with an average of 63 participants and 9 markets per session. Treatments were randomized within each session. In total, each treatment included 17 markets, except for the Seller-Only Discourse treatment, which included 18 markets, resulting in 483 participants across the four experimental conditions. The experiment was implemented using oTree (Chen et al., 2016).

Upon entering the laboratory, participants read the experimental instructions and completed a set of comprehension questions on their computer screens to ensure understanding. The instructions were provided in Spanish; an English translation of the instructions is available in the online Appendix.

After completing the market experiment and the social norms assessment, all participants played a Dictator Game in which they could allocate 30 points either to the charity (GiveDirectly) or themselves. Following this, participants answered a series of closed-ended questions regarding their motivations as sellers and buyers, an open-ended question

allowing them to elaborate on additional reasons for their choices, and a brief demographic questionnaire.

At the end of each session, one period was randomly selected for payment. Experimental points were converted into euros at a rate of 1,000 points to €12.00. On average, participants earned approximately €18, including a €12 show-up fee. Each session lasted about two hours.

Predictions— Our main pre-registered hypotheses (see section E.D.1.f of the pre-analysis plan) concern the fraction of responsible products traded out of all completed trades. Following, Bartling et al. (2024b), we first hypothesize that allowing both sellers and buyers to engage in public discourse (*Full Discourse*) will increase the fraction of responsible products traded relative to the *No Discourse* baseline. Second, we hypothesize that partial discourse by either sellers (*Seller Discourse* treatment) or buyers (*Buyer Discourse* treatment) will also increase the rate of responsible products traded, but by less than in the *Full Discourse* treatment.

3 Results

In what follows, we will follow our pre-analysis plan except when otherwise indicated. The full pre-analysis plan can be found in Appendix E. Figure 1 presents the market share of socially responsible products across different treatment conditions. These shares are calculated based on completed transactions, excluding cases in which buyers made no purchase. The results do not vary substantially if we include all the products offered in a market.

Full Discourse— We begin by investigating the effect of full discourse, replicating the analysis of Bartling et al. (2024b). As demonstrated in Figure 1, introducing full discourse increases the share of socially responsible products from 17% to 25%. This effect is marginally

significant in our pre-registered test ($p = 0.09$, one-sided t -test).² These comparative statics are in line with Bartling et al. (2024b), though the effect size is considerably smaller (8.5pp vs. 28.6pp; $p < 0.05$, two-sided t -test, Table B.2), something we will return to in the next section. It is also worth noting that the *No Discourse* share of socially responsible products in our sample is substantially less than the Swiss sample in Bartling et al. (2024b) (17pp vs. 50pp; $p < 0.01$, two-sided t -test, Table B.2). This suggests that baseline levels of social concern may vary significantly across these two populations.

Partial Discourse— Figure 1 also shows the market share of socially responsible products in the *Seller-Only* and *Buyer-Only* Discourse conditions. The key result is that enabling discourse only on one side of the market is insufficient to shift market behavior. This remains true when we pool both partial discourse conditions to increase statistical power, as can be seen in column (4) of Table 1, which reports the pre-registered regression analysis. Although contrary to our pre-registered expectations, the absence of a partial discourse effect becomes less surprising in light of the relatively small effect observed under full discourse.

Dynamics— To examine whether the treatment effects vary over time, we analyze the dynamic evolution of outcomes across the 24 market rounds. Figure 2 displays the share of socially responsible products traded in each period, disaggregated by treatment condition. We first observe a general decline in the share of responsible products during the initial 7 rounds across all treatments. This pattern also differs from Bartling et al. (2024b), where the share of responsible products is relatively stable and does not display a systematic upward or downward trend over time. Table 2 provides a pre-registered test for dynamics comparing the first and second half of the experiment. The “post” dummy confirms that socially responsible products are observed less often in later periods. The *Full Discourse* treatment

²In four markets, no responsible product was ever traded throughout the session. Excluding these markets from the analysis yields a larger estimated treatment effect of 14pp; $p = 0.019$, one-sided t -test; Figure B.1.

has the highest rates of responsible behavior throughout, though the treatment effect is no longer significant when considering the first or second half of the experiment on its own.

Social Norms— There are also no strong differences in the norm elicitation task (Figure 3). Prior to the first period, participants tend to consider trading the harmful good somewhat socially inappropriate (normalized rating of between -0.24 and -0.36, where a rating of -0.33 equates to “socially inappropriate”). Table 3 shows that discourse does not consistently impact the appropriateness ratings. Unexpectedly, only the *Buyer-Only* treatment reduces the appropriateness of harmful behavior relative to *No Discourse*. Table B.3 shows that this effect (as expected) is driven by the buyers, though it does not survive adjustments for multiple testing. Figure 3 also suggests that market experience reduces the moral qualms associated with trading the harmful good. Table B.1 shows that this change is indeed significant and similar across the four treatments.

Content Analysis— To further delve into potential mechanisms, we analyze the contents of the chat transcripts using OpenAI’s GTP-4o (May 2025 release) and the classification scheme of Bartling et al. (2024b). Note that while we pre-registered doing content analysis, we did not pre-register the exact method to be used; a detailed description of the method is presented in Appendix A. Table B.4 presents the results of this exercise and Table B.5 formally compares the three relevant treatments. The results suggest that participants followed the instructions by talking about responsible and harmful products, fairness, efficiency concerns and prices in all three treatments. In additional analysis presented in the Appendix, we show that the content of the chat predicts market outcomes. At the same time, no clear differences between treatments emerge (Table B.5), which is consistent with the lack of major differences in outcomes between the three discourse treatments.

Prices and Earnings— As a final step, we also examine prices and earnings. Across all four treatments, responsible products are sold for roughly 280 points, whereas harmful products average about 200 points, implying that the 60% of the production cost of the socially responsible products (200 points) is borne by the sellers (see Figure B.2). This is sensible, given that there are more sellers than buyers on the market, and hence buyers have greater bargaining power. Column (1) in Table 4 shows that buyers are price sensitive, being more likely to select the responsible product when its lowest available price falls, or when the cheapest harmful alternative increases in price. Similarly, column (1) in Table 5 indicates that sellers increase the supply of responsible goods as the expected profit advantage of these items—calculated from realized profits in previous periods—improves relative to the harmful alternative. These patterns may help explain why the average price of harmful products decreases over time (Figure B.3), increasing buyer earnings. Falling prices also decrease seller profits over time, though this is partially offset by sellers substituting away from harmful to responsible products, as discussed previously.

4 Discussion

Our main results are (i) a marginally significant effect of full discourse and (ii) no effect of partial discourse on socially responsible behavior. Perhaps the most striking and surprising aspect of these results is that the full discourse effect is only about 25% as large as in Bartling et al. (2024b). In this section, we will examine potential reasons for this discrepancy in greater detail. Some of the analysis in this section uses the data of Bartling et al. (2024b), available from their replication package. None of the analysis is pre-registered, since we did not expect the public discourse effect to differ in our study.

Differences in Social Norms— One possible explanation for the smaller discourse effect observed in our study lies in differences in social norms. In particular, it is possible that

our (Spanish) population has fewer moral qualms about producing harmful goods than the (Swiss) population of Bartling et al. (2024b). This could both decrease socially responsible behavior per se, and also reduce treatment effects if effective discourse requires enough people to object to producing the harmful product in the first place. To investigate potential differences in social norms, we compare our results to the experiment in Bartling et al. (2024b) that also elicited norms both prior and post market interactions (Study 3). We find some evidence that trading the harmful product is seen as more appropriate in our sample both before (-0.24 in our case vs -0.34 in theirs, $p = 0.08$, two-sided t-test) and after (-0.08 in our case vs -0.19 in theirs, $p = 0.08$, two-sided t-test) market interactions. These differences in baseline norms could at least partially explain both the lower rate of socially responsible trades and the reduced effectiveness of public discourse in our setting. At the same time, these differences are small and participants find trading the harmful good to be somewhat socially inappropriate in both studies. Hence, it is likely that other explanations are necessary to fully explain the difference between the two studies.

Differences in Discourse— Another possible explanation relates to differences in the content of public discourse. For this purpose, we applied our content analysis approach (Appendix A) to the data of Bartling et al. (2024b) Study 2, which most closely resembles ours.³ Table B.4 and B.5 show that discourse in the *Full Discourse* treatment in Bartling et al. (2024b) had a greater word count, which is likely driven at least in part by a greater number of participants (11) and a longer time frame (8 minutes). Our chats are also less likely to feature a general discussion or a discussion of prices. More importantly, however, there are no observable differences in the likelihood that a chat features discussions of bad behavior, self-interest, charities, and harmful or responsible product recommendations, and

³In Study 1 the third party took part in the discourse, Study 3 used a different manipulation in their *Full Discourse* treatment.

only minimal differences in discussions concerning efficiency and fairness. It is unlikely that these minor differences can explain the difference in treatment effects between studies.

Differences in Norm Adherence— One final requirement for discourse to be effective is that market participants care about norm adherence in the first place. Tables 4 and 5 examine the drivers of buyer and seller choices respectively. Table 4 (column (2)) shows that buyers in our study choose products based on prices, whereas social norms (elicited before period 1) have no effect. By contrast, for buyers in Bartling et al. (2024b) norms are highly important and prices matter less (column (4)). Table 5 shows a similar pattern for sellers, who care more about profit differences between products and less about the norm than in Bartling et al. (2024b).⁴ In other words, in contrast to Bartling et al. (2024b), our buyers and sellers appear to prioritize payments over norm adherence. This may explain both the lower rate of socially responsible behavior, and the smaller public discourse effects we observe in our study.

Other Differences— Finally, it is worth noting that our study differed from Bartling et al. (2024b) in several other ways. This includes the charity we used, using smaller markets (7 rather than 11 participants), the use of Otree rather than Ztree, and adding more visuals to the instructions. We also used a slightly shorter discourse time (5 instead of 8 minutes) because of a pilot study in which we found that the vast majority of discourse happened in the first five minutes. Hence, our study is not a perfect replication. Nevertheless, these changes are small, and we at least did not expect them to affect the results in a meaningful way. Instead, we attribute the lack of a public discourse effect in our study to differences in social norms and to participants prioritizing profits over norm adherence in our study.

⁴We conduct these analyses using Study 3 from Bartling et al. (2024b), as it is the only study in which social norms were elicited prior to market interaction. However, we obtain similar results by instead using the post-market norm and the data of Study 2.

5 Conclusion

This study investigates whether public discourse can promote socially responsible behavior in markets. Our results show that when both sellers and buyers are able to converse, the share of responsible trades increases by about eight percentage points relative to the no-discourse baseline. By contrast, when discourse is confined to only one side of the market, no significant effects on behavior is observed. A content analysis of chat transcripts suggests that limiting discourse reduces the volume of discussion, but not the thematic emphasis of arguments.

These findings replicate the main qualitative insight of Bartling et al. (2024b)—that open discourse can foster more responsible market behavior—but with a substantially smaller effect size. Whereas Bartling et al. (2024b) report a 29 percentage point increase, our *Full Discourse treatment* is much smaller (8pp). We present evidence that suggests that these differences may be caused both by differences in underlying social norms and the fact that our participants, in contrast to Bartling et al. (2024b), appear to prioritize profits over norm adherence.

These results add worthwhile context to previous work on socially responsible markets. Previous work suggests that interventions such as public discourse may effectively increase responsible behavior. Yet, what our results show is that the effectiveness of such interventions may depend on the extent to which market participants are willing to prioritize norms over profit in the first place. This matters when considering the external validity of such interventions. Our results also point to the potential importance of subject pool effects in experimental economics, and potential caveats for studies relying on only a single population.

Future work should further explore the conditions of discourse-based interventions. One important way is to replicate this design in other cultural contexts, where baseline norms and sensitivities may differ. Indeed, it is plausible that partial discourse may be more effective in settings where market participants care about norm adherence. Finally, extending the

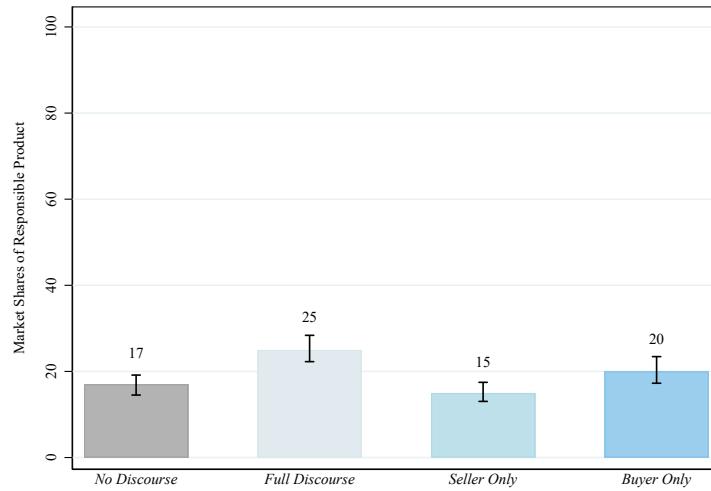
analysis to field or online marketplace settings would help assess the external validity of the mechanism and its potential as a cost-effective policy tool. Together, such efforts would contribute to a clearer understanding of when and how public discourse can reliably promote socially responsible behavior in markets.

References

- Agranov, Marina and Chloe Tergiman**, “Communication in multilateral bargaining,” *Journal of Public Economics*, 2014, *118*, 75–85.
- Andreoni, James and Justin M Rao**, “The power of asking: How communication affects selfishness, empathy, and altruism,” *Journal of public economics*, 2011, *95* (7-8), 513–520.
- Baron, David P**, “Private politics, corporate social responsibility, and integrated strategy,” *Journal of Economics & Management Strategy*, 2001, *10* (1), 7–45.
- Bartling, Björn, Roberto A Weber, and Lan Yao**, “Do markets erode social responsibility?,” *The Quarterly Journal of Economics*, 2015, *130* (1), 219–266.
- , **Vanessa Valero, and Roberto A Weber**, “The causal effect of income on market social responsibility,” Technical Report, Working Paper 2024.
- , – , **and Roberto Weber**, “On the scope of externalities in experimental markets,” *Experimental Economics*, 2019, *22* (3), 610–624.
- , – , **Roberto A Weber, and Lan Yao**, “Public discourse and socially responsible market behavior,” *American Economic Review*, 2024, *114* (10), 3041–3074.
- Bénabou, Roland and Jean Tirole**, “Individual and corporate social responsibility,” *Economica*, 2010, *77* (305), 1–19.
- Bochet, Olivier, Talbot Page, and Louis Putterman**, “Communication and punishment in voluntary contribution experiments,” *Journal of Economic Behavior & Organization*, 2006, *60* (1), 11–26.
- Charness, Gary and Martin Dufwenberg**, “Promises and partnership,” *Econometrica*, 2006, *74* (6), 1579–1601.
- Chen, Daniel L, Martin Schonger, and Chris Wickens**, “oTree—An open-source platform for laboratory, online, and field experiments,” *Journal of Behavioral and Experimental Finance*, 2016, *9*, 88–97.
- Crane, Andrew, Dirk Matten, Sarah Glozer, and Laura J Spence**, *Business ethics: Managing corporate citizenship and sustainability in the age of globalization*, Oxford university press, 2019.
- Dawes, Robyn M, Jeanne McTavish, and Harriet Shaklee**, “Behavior, communication, and assumptions about other people’s behavior in a commons dilemma situation.,” *Journal of Personality and Social Psychology*, 1977, *35* (1), 1.

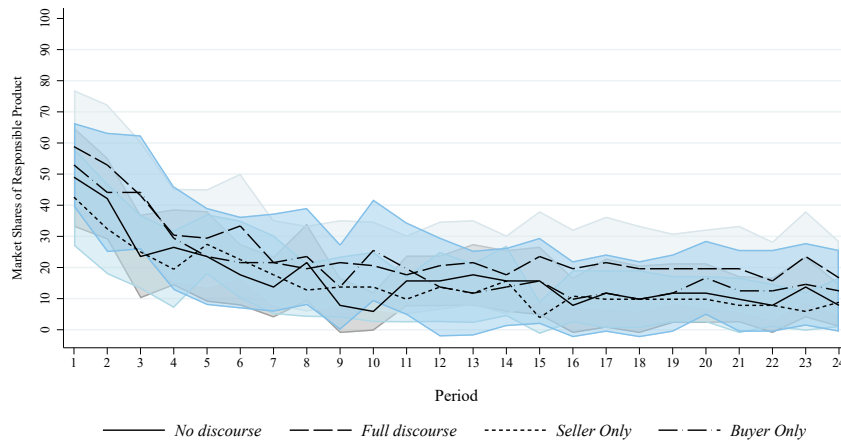
- Dufwenberg, Martin, Olof Johansson-Stenman, Michael Kirchler, Florian Lindner, and Rene Schwaiger**, “Mean markets or kind commerce?,” *Journal of Public Economics*, 2022, *209*, 104648.
- , **Paul Heidhues, Georg Kirchsteiger, Frank Riedel, and Joel Sobel**, “Other-regarding preferences in general equilibrium,” *The Review of Economic Studies*, 2011, *78* (2), 613–639.
- Krupka, Erin L and Roberto A Weber**, “Identifying social norms using coordination games: Why does dictator game sharing vary?,” *Journal of the European Economic Association*, 2013, *11* (3), 495–524.
- Micheletti, Michele**, *Political virtue and shopping: Individuals, consumerism, and collective action*, Palgrave Macmillan, New York, 2003.
- Rode, Julian, Robin M Hogarth, and Marc Le Menestrel**, “Ethical differentiation and market behavior: An experimental approach,” *Journal of Economic Behavior & Organization*, 2008, *66* (2), 265–280.
- Smith, N Craig**, “Consumers as drivers of corporate social responsibility,” *Mimeo*, 2008.

Figure 1 Market share of socially responsible products



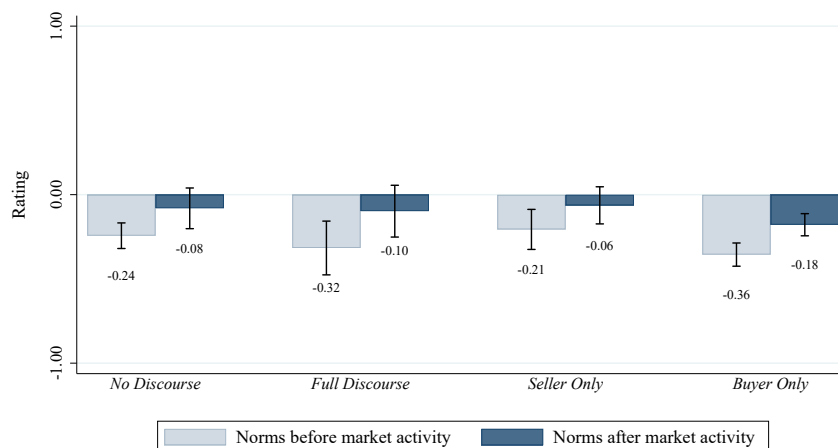
Notes: This figure reports the market share of socially responsible products as well as the 95% confidence intervals calculated at the market level per treatment type. We exclude the cases where a transaction was not completed.

Figure 2 Market share of socially responsible products over the periods



Notes: This figure reports the market share of socially responsible products as well as the 95% confidence intervals calculated at the market level per treatment type over the 24 periods. We exclude the cases where a transaction was not completed.

Figure 3 Effect of public discourse on social norms



Notes: This figure reports the mean rating of the appropriateness of trading the harmful product. Norms before market activity refers to social norms measured prior to interacting in the market; Norms after market activity refers to social norms measured after market interaction. Ratings are elicited in an incentive-compatible way following Krupka and Weber (2013): “very socially appropriate” = 1, “somewhat socially appropriate” = 1/3, “somewhat socially inappropriate” = -1/3, and “very socially inappropriate” = -1. Bars display 95% confidence intervals at the market level.

Table 1 Effect of discourse on socially responsible behavior

	Dependent Variable: Market Share of Responsible Product				
	(1)	(2)	(3)	(4)	(5)
Full discourse	0.0846*		0.1008*	0.0846*	
	(0.0627)		(0.0676)	(0.0627)	
Seller only	-0.0161	-0.1008*			
	(0.0427)	(0.0676)			
Buyer only	0.0341	-0.0505	0.0503		
	(0.0569)	(0.0774)	(0.0623)		
No discourse		-0.0846*	0.0161		-0.0846*
		(0.0627)	(0.0427)		(0.0627)
Partial discourse				0.0088	-0.0759
				(0.0397)	(0.0657)
Constant	0.1681***	0.2527***	0.1519***	0.1681***	0.2527***
	(0.0243)	(0.0578)	(0.0351)	(0.0243)	(0.0578)
No. of observations	4,768	4,768	4,768	4,768	4,768
No. of markets	69	69	69	69	69
R ²	0.010	0.010	0.010	0.008	0.008
Reference treatment	No discourse	Full discourse	Seller only	No discourse	Full discourse
P-value type	one-sided	one-sided	mixed [†]	one-sided	one-sided

Notes: The dependent variable takes the value 1 if a buyer purchased the socially responsible product and 0 otherwise; transactions with no purchase are excluded. Robust standard errors are clustered at the market level. Stars reflect one-sided thresholds unless noted: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. [†] Column 3 uses one-sided tests for coefficients referenced to Full discourse and No discourse (per the pre-analysis direction) but a *two-sided* test for *Buyer only*.

Table 2 Dynamic effects of discourse on socially responsible behavior

	Dependent Variable: Market Share of Responsible Product				
	(1)	(2)	(3)	(4)	(5)
Post	-0.1014*** (0.0367)	-0.1052*** (0.0365)	-0.1153*** (0.0171)	-0.1014*** (0.0367)	-0.1052*** (0.0364)
Full discourse	0.0861 (0.0695)		0.0962 (0.0741)	0.0861 (0.0695)	
Full discourse \times Post	-0.0037 (0.0517)		0.0102 (0.0403)	-0.0037 (0.0517)	
Seller only	-0.0101 (0.0506)	-0.0962 (0.0741)			
Seller only \times Post	-0.0139 (0.0405)	-0.0102 (0.0403)			
Buyer only	0.0586 (0.0706)	-0.0275 (0.0890)	0.0687 (0.0752)		
Buyer only \times Post	-0.0505 (0.0725)	-0.0467 (0.0724)	-0.0366 (0.0649)		
No discourse		-0.0861 (0.0695)	0.0101 (0.0506)		-0.0861 (0.0695)
No discourse \times Post		0.0037 (0.0517)	0.0139 (0.0405)		0.0037 (0.0517)
Partial discourse				0.0239 (0.0488)	-0.0622 (0.0729)
Partial discourse \times Post				-0.0320 (0.0489)	-0.0282 (0.0487)
Constant	0.2194*** (0.0307)	0.3055*** (0.0623)	0.2093*** (0.0402)	0.2194*** (0.0307)	0.3055*** (0.0623)
No. of observations	4,768	4,768	4,768	4,768	4,768
No. of markets	69	69	69	69	69
R ²	0.033	0.033	0.033	0.030	0.030
Reference treatment	No discourse	Full discourse	Seller only	No discourse	Full discourse
P-value type	one-sided	one-sided	mixed [†]	one-sided	one-sided

Notes: The dependent variable equals 1 if a buyer purchased the socially responsible product and 0 otherwise; transactions with no purchase are excluded. The reference treatment is listed per column together with the *Pre* period (rounds 1–12). *Post* is an indicator for rounds 13–24; interaction terms are relative to this *Pre* baseline. Robust standard errors are clustered at the market level. Stars denote one-sided significance thresholds unless noted: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. [†] Column 3 uses one-sided tests for coefficients referenced to *Full discourse* and *No discourse* (per the pre-registered direction), but a *two-sided* test for *Buyer only*.

Table 3 Effect of discourse on norm elicitation

	Dependent Variable: Rating			
	Start		End	
	OLS (1)	Probit (2)	OLS (3)	Probit (4)
Full discourse	-0.0728 (0.0883)	-0.1756 (0.2108)	-0.0168 (0.0978)	-0.0335 (0.2069)
Seller only	0.0373 (0.0706)	0.0894 (0.1606)	0.0177 (0.0818)	0.0381 (0.1730)
Buyer only	-0.1120** (0.0513)	-0.2444** (0.1238)	-0.0973* (0.0688)	-0.1969* (0.1476)
Constant	-0.2437*** (0.0379)		-0.0812* (0.0604)	
No. of observations	483	483	476	476
No. of markets	69	69	68	68
R ²	0.015		0.007	
Reference treatment	No discourse	No discourse	No discourse	No discourse
P-value type	one-sided	one-sided	one-sided	one-sided

Notes: The dependent variable is the incentivized social appropriateness rating of trading the harmful product, measured before trading (Start) and after trading (End). Ratings are elicited in an incentive-compatible way following Krupka and Weber (2013), where the scale takes the values: very socially appropriate = 1, somewhat socially appropriate = 1/3, somewhat socially inappropriate = -1/3, and very socially inappropriate = -1. Robust standard errors are clustered at the market level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4 Panel regressions of responsible buyer product choice

	Dependent Variable: Bought Responsible Product			
	Our Study		Bartling et al (2024)	
	Without norm (1)	With norm (2)	Without norm (3)	With norm (4)
Lowest price of responsible product	-0.0019*** (0.0002)	-0.0019*** (0.0002)	-0.0010** (0.0004)	-0.0010** (0.0004)
Lowest price of harmful product	0.0019*** (0.0002)	0.0019*** (0.0002)	0.0012*** (0.0003)	0.0012*** (0.0003)
Pre-market norm		-0.0077 (0.0499)		-0.1289** (0.0548)
Constant	0.4960*** (0.0836)	0.4839*** (0.0876)	0.6026*** (0.1361)	0.5754*** (0.1222)
No. of observations	1,818	1,766	4,141	4,141
No. of markets	65	65	56	56
R ²	0.098	0.099	0.016	0.027

Notes: The dependent variable takes the value of 1 if a buyer purchased a responsible product and 0 otherwise. Lowest price of responsible product and Lowest price of harmful product refer to the prices of products available to the buyer. Pre-market norm is the incentivized norm elicited before any market interactions. The model omits the cases in which a buyer made no product purchase and cases in which either only responsible or harmful products were available to a buyer. The first two columns use data from our experiment, columns 3 and 4 use data from Bartling et al. (2024, Study 3, No Discourse and Discourse (Neutral) treatments). We fix the scale of Bartling et al. (2024) to be similar to ours. Robust standard errors are clustered at the market level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5 Panel regressions of responsible seller product decisions

	Dependent Variable: Offer Responsible Product			
	Our Study		Bartling et al (2024)	
	Without norm (1)	With norm (2)	Without norm (3)	With norm (4)
Expected responsible product profit premium	0.0007*** (0.0001)	0.0007*** (0.0001)	0.0005*** (0.0001)	0.0005*** (0.0001)
Pre-market norm		-0.1040** (0.0433)		-0.2286*** (0.0455)
Constant	0.3761*** (0.0246)	0.3408*** (0.0267)	0.5721*** (0.0226)	0.4954*** (0.0253)
No. of observations	2,972	2,972	6,282	6,282
No. of markets	65	65	55	55
R ²	0.026	0.037	0.008	0.056

Notes: The dependent variable in all models is a binary variable taking the value of 1 if a seller offered a responsible product and 0 otherwise. The variable Expected responsible product profit premium measures the average realized profit difference between offering a responsible product and offering a harmful product in the preceding period. Pre-market norm is the incentivized norm elicited before any market interactions. Note that if an offer is not accepted, the seller's profit equals zero. The first two columns use data from our experiment, columns 3 and 4 use data from Bartling et al. (2024, Study 3, No Discourse and Discourse (Neutral) treatments). We fix the scale of Bartling et al. (2024) to be similar to ours. Robust standard errors are clustered at the market level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Online Appendix for “Cultural Foundations of Socially Responsible Behavior”

A Content Analysis

We conduct a chat-level content analysis of the transcripts using a large language model (LLM; OpenAI GPT-4o, May 2025 release).⁵ For each chat session j , we merge all individual messages into a single text document D_j , while preserving speaker labels and time stamps to keep track of the dialogue flow. In other words, the LLM analyzes the entire chat transcript as one unit rather than evaluating each message separately. The model is then asked to score D_j against a pre-specified set of K thematic categories—developed from the coding scheme used in Bartling et al. (2024b)—such as RESPONSIBLE PRODUCT RECOMMENDATION, PRICE DISCUSSION, and FAIRNESS. The output is a vector

$$\mathbf{s}_j = (s_{1j}, \dots, s_{Kj}), \quad s_{kj} \in [0, 1],$$

where s_{kj} is the model’s estimated relevance—or semantic coverage—of category k in the entire transcript. A score of 0 indicates no detectable presence of the theme, whereas 1 denotes maximal alignment.

To allow us to compare our results to Bartling et al. (2024b), we apply the identical classification protocol and thematic taxonomy to the chat transcripts from our experiment and to the publicly available transcripts of Bartling et al. (2024b). For every chat j in each data set we obtain the vector $\mathbf{s}_j = (s_{1j}, \dots, s_{Kj})$ of category scores described above. We then compare the results of s_{kj} across treatments and studies to assess whether systematic

⁵We fixed the sampling parameters (`temperature=0`, `top_p=1`) so that the model returns a deterministic answer to a given prompt, an essential requirement for replicable analysis.

differences in chat content can account for (a) the treatment effects in our study and (b) the difference between our study and Bartling et al. (2024b).

We next examine whether chat content predicts both subsequent trading decisions and participants’ beliefs about the appropriateness of trading the harmful product, thereby gauging the extent to which public discourse can influence market outcomes. Specifically, we relate the chat-level LLM scores for recommendations favoring the responsible product and those favoring the harmful product to the average share of responsible trades executed in each market and the corresponding norm score.

Table B.7 shows that markets in which the chat more frequently recommends trading the responsible product exhibit a higher average share of socially responsible trades. Conversely, increased discussion that favors the harmful product is associated with a lower share of responsible trades. These trading patterns align with the observed shifts in normative beliefs. Chats that emphasize the responsible product yield lower appropriateness ratings for trading the harmful product, whereas chats that advocate the harmful product yield higher appropriateness ratings. To gauge the net directional tone of each chat, we construct (i) a continuous difference score—the LLM probability assigned to recommending the responsible product minus the probability assigned to recommending the harmful product—and (ii) a median-split indicator based on this difference. Both measures reproduce the same qualitative relationships: a more pro-responsible tone predicts a larger share of responsible trades and stricter norms against the harmful product.

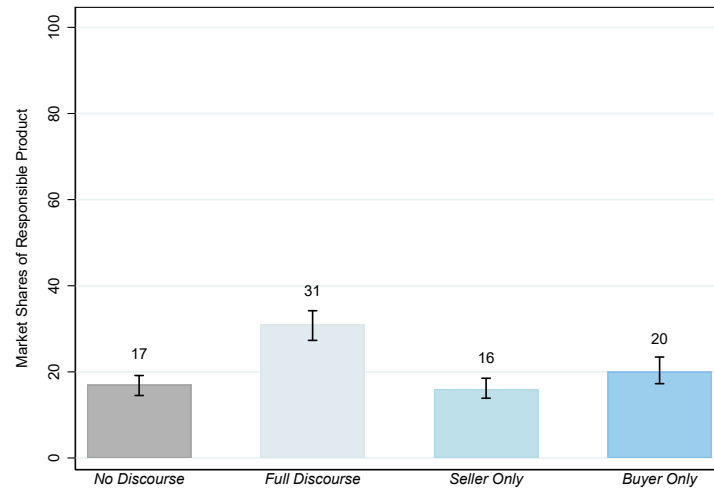
We next disaggregate these relationships by treatment. Table B.8 shows that the *difference* score positively predicts the share of responsible trades in all three treatments, though this effect is not significant in the *Buyer-Only* treatment. Its effect on the norm score, however, is heterogeneous. A more pro-responsible chat is associated with stricter norms against the harmful product in both the *Full Discourse* and *Seller-Only* conditions, but no such relationship emerges in the *Buyer-Only* condition. These findings are robust to

the alternative chat-content measures introduced earlier; all specifications yield qualitatively identical treatment-specific patterns.

In sum, the evidence indicates that—even when communication is confined to a single side of the market—a more prosocial tone in the chat is still associated with improved aggregate outcomes, namely a higher market share of socially responsible products. Note that these estimates are correlational rather than causal, since it is likely that markets with responsible individuals would both have different discourses and different market behavior.

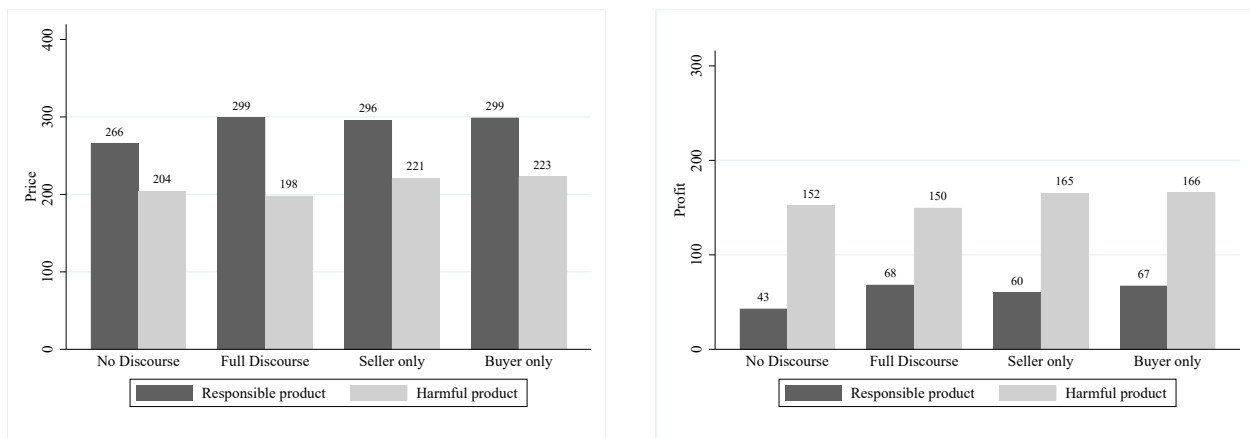
B Figures and Tables

Figure B.1 Market share of socially responsible products (without outliers)



Notes: This figure reports the market share of socially responsible products as well as the 95% confidence intervals calculated at the market level per treatment type. We exclude the cases where a transaction was not completed. The figure also excludes four markets in which no responsible product was traded at any point during the session.

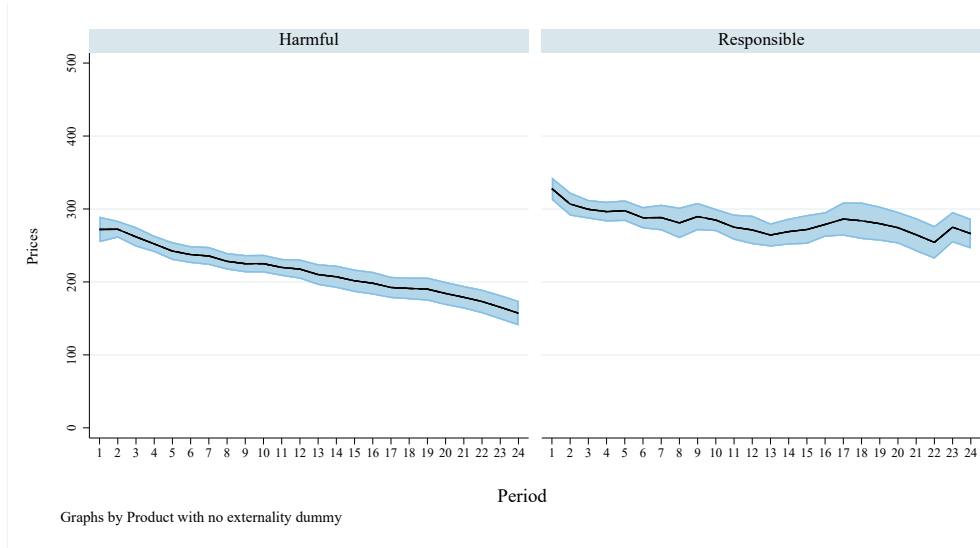
Figure B.2 Prices and profits by product type and treatment group
 (a) Prices (b) Profit



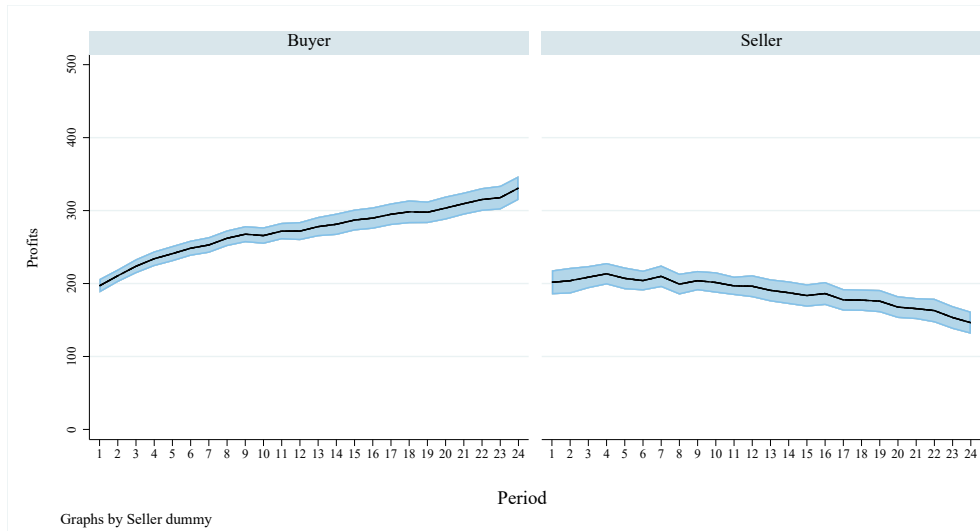
Notes: These figures present the mean transaction price and the mean seller profit for each product type, disaggregated by treatment. Price averages are calculated only for completed trades and exclude observations in which the buyer made no purchase. Seller profit equals the transaction price minus the production cost when the product is sold and is set to zero if the offer is not accepted. product.

Figure B.3 Dynamic evolution of prices and profits by product type

(a) Prices

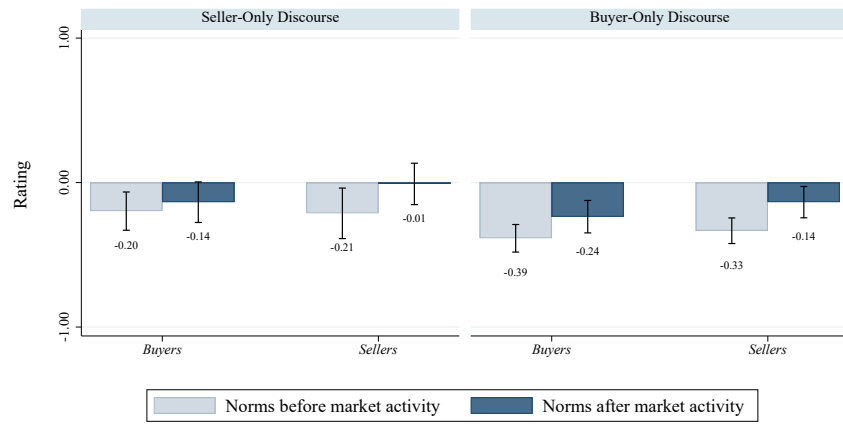


(b) Profit



Notes: These figures present the mean transaction price for each product type and the mean profit for each player type over the 24 periods. Price averages are calculated only for completed trades and exclude observations in which the buyer made no purchase. Seller profit equals the transaction price minus the production cost when the product is sold and is set to zero if the offer is not accepted. Buyer profit equals the 500 points value of the product minus the transaction price when a product is bought and zero if no product is bought.

Figure B.4 Effect of public discourse on social norms: seller- and buyer-only treatments



Graphs by (mean) treatment

Notes: This figure reports the mean rating of the appropriateness of trading the harmful product. Norms before market activity refers to social norms measured prior to interacting in the market; Norms after market activity refers to social norms measured after market interaction. Ratings are elicited in an incentive-compatible way following Krupka and Weber (2013): “very socially appropriate” = 1, “somewhat socially appropriate” = 1/3, “somewhat socially inappropriate” = -1/3, and “very socially inappropriate” = -1. Bars display 95% confidence intervals at the market level.

Table B.1 Effect of Discourse on norm elicitation by player type and pre-/post social norm

	All		Buyers		Sellers	
	OLS (1)	Probit (2)	OLS (3)	Probit (4)	OLS (5)	Probit (6)
Full discourse	-0.0728 (0.0884)	-0.1713 (0.2066)	-0.0784 (0.1198)	-0.2014 (0.2829)	-0.0686 (0.0960)	-0.1466 (0.2244)
Seller only	0.0373 (0.0706)	0.0872 (0.1583)	-0.0080 (0.0912)	-0.0101 (0.2129)	0.0714 (0.1002)	0.1537 (0.2242)
Buyer only	-0.1120** (0.0513)	-0.2434** (0.1204)	-0.1961** (0.0785)	-0.4656** (0.1891)	-0.0490 (0.0656)	-0.0889 (0.1522)
Ex post norms	0.1625*** (0.0542)	0.3573*** (0.1181)	0.0915 (0.0753)	0.2034 (0.1719)	0.2157** (0.0886)	0.4705** (0.1897)
Full discourse × Ex post norms	0.0560 (0.0853)	0.1372 (0.1925)	0.0523 (0.1095)	0.1505 (0.2533)	0.0588 (0.1190)	0.1216 (0.2591)
Seller only × Ex post norms	-0.0196 (0.0683)	-0.0479 (0.1459)	-0.0298 (0.0969)	-0.0777 (0.2194)	-0.0120 (0.1088)	-0.0258 (0.2302)
Buyer only × Ex post norms	0.0147 (0.0714)	0.0466 (0.1580)	0.0580 (0.0948)	0.1677 (0.2202)	-0.0178 (0.1102)	-0.0454 (0.2369)
Constant	-0.2437*** (0.0380)		-0.1895*** (0.0621)		-0.2843*** (0.0482)	
No. of observations	959	959	411	411	548	548
No. of markets	69	69	69	69	69	69
R ²	0.041		0.033		0.054	
Reference treatment	No discourse Pre norm	No discourse Pre norm	No discourse Pre norm	No discourse Pre norm	No discourse Pre norm	No discourse Pre norm
P-value type	one-sided	one-sided	one-sided	one-sided	one-sided	one-sided

Notes: The dependent variable is the incentivized social appropriateness rating of trading the harmful product, measured before trading (Start) and after trading (End). Ratings are elicited in an incentive-compatible way following Krupka and Weber (2013), where the scale takes the values: very socially appropriate = 1, somewhat socially appropriate = 1/3, somewhat socially inappropriate = -1/3, and very socially inappropriate = -1. Robust standard errors are clustered at the market level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.2 Effect of discourse on socially responsible behavior across studies

	Market Share of Responsible Product	
	(1)	(2)
Full discourse × Bartling	0.2018** (0.0950)	
Full discourse × Our study		-0.2018** (0.0950)
Full discourse	0.0846* (0.0627)	0.2864*** (0.0713)
Bartling study	0.3309*** (0.0573)	
Our study		-0.3309*** (0.0573)
Constant	0.1681*** (0.0243)	0.4989*** (0.0519)
No. of observations	6,133	6,133
No. of markets	66	66
R ²	0.231	0.231
Reference treatment	No discourse (Our Study)	No discourse (Bartling)
P-value type	one-sided	one-sided

Notes: The dependent variable equals 1 if a buyer purchased the socially responsible product and 0 otherwise. Transactions in which the buyer made no purchase are excluded. Model 1 estimates treatment effects relative to the No Discourse condition in our study, whereas Model 2 uses the No Discourse condition from Bartling et al. (2024b) as the reference treatment. This choice of reference treatments facilitates direct comparisons of treatment effects both within our study and across studies. Robust standard errors are clustered at the market level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.3 Effect of discourse on norm elicitation by player type

	Dependent Variable: Rating							
	Sellers				Buyers			
	Start		End		Start		End	
	OLS (1)	Probit (2)	OLS (3)	Probit (4)	OLS (5)	Probit (6)	OLS (7)	Probit (8)
Full discourse	-0.0686 (0.0959)	-0.1457 (0.2282)	-0.0098 (0.1185)	-0.0229 (0.2445)	-0.0784 (0.1197)	-0.2137 (0.2895)	-0.0261 (0.1156)	-0.0530 (0.2566)
Seller only	0.0714 (0.1001)	0.1530 (0.2268)	0.0594 (0.1026)	0.1253 (0.2094)	-0.0080 (0.0910)	-0.0051 (0.2183)	-0.0378 (0.1088)	-0.0864 (0.2436)
Buyer only	-0.0490 (0.0655)	-0.0847 (0.1551)	-0.0668 (0.0910)	-0.1341 (0.1884)	-0.1961*** (0.0784)	-0.4766*** (0.1940)	-0.1381* (0.1001)	-0.2989* (0.2259)
Constant	-0.2843*** (0.0482)		-0.0686 (0.0731)		-0.1895*** (0.0620)		-0.0980 (0.0828)	
No. of observations	276	276	272	272	207	207	204	204
No. of markets	69	69	68	68	69	69	68	68
R ²	0.012		0.007		0.029		0.011	
Reference treatment	No discourse	No discourse	No discourse	No discourse	No discourse	No discourse	No discourse	No discourse
P-value type	one-sided	one-sided	one-sided	one-sided	one-sided	one-sided	one-sided	one-sided

Notes: The dependent variable is the incentivized social appropriateness rating of trading the harmful product, measured before trading (Start) and after trading (End). Ratings are elicited in an incentive-compatible way following Krupka and Weber (2013), where the scale takes the values: very socially appropriate = 1, somewhat socially appropriate = 1/3, somewhat socially inappropriate = -1/3, and very socially inappropriate = -1. Robust standard errors are clustered at the market level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.4 Content analysis of chats

Variable	Summary Statistics by Treatment									
	Our Study						Bartling et al. (2024)			
	Full Discourse		Seller Only		Buyer Only		Full Discourse		Experienced	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Responsible Product Recommendation	0.447	(0.310)	0.417	(0.322)	0.382	(0.288)	0.488	(0.322)	0.694	(0.246)
Harmful Product Recommendation	0.341	(0.276)	0.294	(0.271)	0.365	(0.271)	0.319	(0.293)	0.188	(0.136)
Bad Action	0.135	(0.117)	0.139	(0.138)	0.153	(0.197)	0.119	(0.075)	0.125	(0.086)
Price Discussion	0.376	(0.303)	0.350	(0.287)	0.200	(0.257)	0.650	(0.190)	0.763	(0.175)
Fairness	0.606	(0.185)	0.539	(0.250)	0.524	(0.282)	0.725	(0.177)	0.763	(0.120)
Efficiency	0.406	(0.192)	0.267	(0.209)	0.276	(0.205)	0.500	(0.121)	0.569	(0.101)
Self Interest	0.529	(0.214)	0.544	(0.192)	0.476	(0.295)	0.475	(0.153)	0.487	(0.175)
Mention GiveDirectly	0.000	(0.000)	0.000	(0.000)	0.012	(0.049)	0.000	(0.000)	0.000	(0.000)
Market Questions	0.065	(0.111)	0.022	(0.055)	0.047	(0.094)	0.156	(0.167)	0.169	(0.149)
General Discussion	0.241	(0.142)	0.217	(0.138)	0.118	(0.155)	0.406	(0.118)	0.350	(0.121)
Unclear	0.065	(0.049)	0.072	(0.057)	0.094	(0.241)	0.088	(0.034)	0.081	(0.040)
Word count	273.824	(102.678)	168.167	(61.102)	110.588	(59.761)	572.375	(136.505)	651.938	(174.321)
Character count	1903.647	(671.820)	1197.056	(410.318)	761.412	(401.025)	3478.438	(773.338)	3952.875	(1074.304)
No. of Chats	17		18		17		16		16	

Notes: This table presents means and standard deviations of key variables, grouped by treatment. The variables are measured as the probability (ranging from 0 to 1) that a discussion belongs to each category, based on its overall content. Higher values indicate stronger relevance. **Responsible Product Recommendation:** The discussion recommends trading products without an effect on the donation or considers it inappropriate to trade products with a reduction in the donation. **Harmful Product Recommendation:** The discussion recommends trading products with a reduction on the donation or considers it appropriate to trade products with a reduction in the donation. **Bad Action:** The discussion suggests that more than 50% of the products traded will be with a reduction in the donation. **Price Discussion:** The discussion focuses on the prices of traded products. **Fairness:** The discussion includes considerations of morality, justice, or concerns related to poverty. **Efficiency:** The discussion emphasizes efficiency or total income maximization as a central topic. **Self Interest:** The discussion primarily argues from a self-interest perspective or prioritizes profit maximization. **Mention GiveDirectly:** The discussion includes references to the charity GiveDirectly or its goals. **Market Questions:** The discussion involves questions about how to participate in the market game. **General Discussion:** The discussion revolves around the experiment but does not include clear recommendations regarding trading decisions. **Unclear:** The discussion does not fit into a clear category or contains ambiguous content. **Word Count:** The total number of words in the discussion. **Character Count:** The total number of characters in the discussion. Standard deviations are in parentheses. Standard deviations are in parentheses.

Table B.5 Differences in chats between studies

Variable	Dependent variables										
	Responsible Product Recommendation (1)	Harmful Product Recommendation (2)	Bad Action (3)	Price Discussion (4)	Fairness (5)	Efficiency (6)	Self Interest (7)	Mention GiveDirectly (8)	Market Question (9)	General Discussion (10)	Unclear (11)
Full discourse X Bartling	0.0404 (0.1102)	-0.0224 (0.0991)	-0.0165 (0.0340)	0.2735*** (0.0875)	0.1191* (0.0630)	0.0941* (0.0555)	-0.0544 (0.0645)	-0.0000 (0.0000)	0.0915* (0.0497)	0.1651*** (0.0453)	0.0228 (0.0147)
Buyer only	-0.0647 (0.1027)	0.0235 (0.0940)	0.0176 (0.0556)	-0.1765* (0.0965)	-0.0824 (0.0818)	-0.1294* (0.0681)	-0.0529 (0.0884)	0.0118 (0.0118)	-0.0176 (0.0354)	-0.1235** (0.0510)	0.0294 (0.0597)
Seller only	-0.0304 (0.1071)	-0.0467 (0.0927)	0.0036 (0.0432)	-0.0265 (0.1001)	-0.0670 (0.0742)	-0.1392** (0.0678)	0.0150 (0.0690)	-0.0000 (0.0000)	-0.0425 (0.0300)	-0.0245 (0.0474)	0.0075 (0.0181)
Experienced	0.2467** (0.0972)	-0.1537** (0.0751)	-0.0103 (0.0355)	0.3860*** (0.0855)	0.1566*** (0.0541)	0.1629*** (0.0530)	-0.0419 (0.0679)	-0.0000 (0.0000)	0.1040** (0.0460)	0.1088** (0.0458)	0.0165 (0.0156)
Constant	0.4471*** (0.0753)	0.3412*** (0.0670)	0.1353*** (0.0284)	0.3765*** (0.0736)	0.6059*** (0.0450)	0.4059*** (0.0466)	0.5294*** (0.0520)	0.0000 (0.0000)	0.0647** (0.0270)	0.2412*** (0.0344)	0.0647*** (0.0120)
No. of markets	84	84	84	84	84	84	84	84	84	84	84
R ²	0.122	0.056	0.008	0.417	0.179	0.335	0.020	0.047	0.204	0.369	0.009

Notes: The reference treatment refers to the Full Discourse in our study. The variables are measured as the probability (0–1) that a discussion belongs to each category, based on overall content. Higher values indicate stronger relevance. **Harmful Product Recommendation:** The discussion recommends trading products with a reduction on the donation or considers it appropriate to trade products with a reduction in the donation. **Bad Action:** The discussion suggests that more than 50% of the products traded will be with a reduction in the donation. **Price Discussion:** The discussion focuses on the prices of traded products. **Fairness:** The discussion includes considerations of morality, justice, or concerns related to poverty. **Efficiency:** The discussion emphasizes efficiency or total income maximization as a central topic. **Self Interest:** The discussion primarily argues from a self-interest perspective or prioritizes profit maximization. **Mention GiveDirectly:** The discussion includes references to the charity GiveDirectly or its goals. **Market Questions:** The discussion involves questions about how to participate in the market game. **General Discussion:** The discussion revolves around the experiment but does not include clear recommendations regarding trading decisions. **Unclear:** The discussion does not fit into a clear category or contains ambiguous content. Robust standard errors are clustered at the market level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.6 Differences across chats

Variable	Dependent variables										
	Responsible Product Recommendation (1)	Harmful Product Recommendation (2)	Bad Action (3)	Price Discussion (4)	Fairness (5)	Efficiency (6)	Self Interest (7)	Mention GiveDirectly (8)	Market Question (9)	General Discussion (10)	Unclear (11)
Full discourse	0.0647 (0.1026)	-0.0235 (0.0939)	-0.0176 (0.0556)	0.1765* (0.0964)	0.0824 (0.0818)	0.1294* (0.0680)	0.0529 (0.0884)	-0.0118 (0.0118)	0.0176 (0.0354)	0.1235** (0.0509)	-0.0294 (0.0596)
Seller only	0.0343 (0.1032)	-0.0703 (0.0918)	-0.0141 (0.0578)	0.1500 (0.0922)	0.0154 (0.0903)	-0.0098 (0.0699)	0.0680 (0.0846)	-0.0118 (0.0118)	-0.0248 (0.0263)	0.0990* (0.0498)	-0.0219 (0.0600)
Constant	0.3824*** (0.0697)	0.3647*** (0.0658)	0.1529*** (0.0478)	0.2000*** (0.0624)	0.5235*** (0.0683)	0.2765*** (0.0496)	0.4765*** (0.0715)	0.0118 (0.0118)	0.0471** (0.0229)	0.1176*** (0.0376)	0.0941 (0.0584)
No. of markets	52	52	52	52	52	52	52	52	52	52	52
R ²	0.008	0.012	0.003	0.073	0.022	0.094	0.016	0.040	0.039	0.124	0.008

Notes: The reference treatment refers to the Buyer Discourse treatment. The variables are measured as the probability (ranging from 0 to 1) that a discussion belongs to each category, based on its overall content. Higher values indicate stronger relevance. **Harmful Product Recommendation:** The discussion recommends trading products with a reduction on the donation or considers it appropriate to trade products with a reduction in the donation. **Bad Action:** The discussion suggests that more than 50% of the products traded will be with a reduction in the donation. **Price Discussion:** The discussion focuses on the prices of traded products. **Fairness:** The discussion includes considerations of morality, justice, or concerns related to poverty. **Efficiency:** The discussion emphasizes efficiency or total income maximization as a central topic. **Self Interest:** The discussion primarily argues from a self-interest perspective or prioritizes profit maximization. **Mention GiveDirectly:** The discussion includes references to the charity GiveDirectly or its goals. **Market Questions:** The discussion involves questions about how to participate in the market game. **General Discussion:** The discussion revolves around the experiment but does not include clear recommendations regarding trading decisions. **Unclear:** The discussion does not fit into a clear category or contains ambiguous content. Robust standard errors are clustered at the market level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.7 Chat content and market behavior

Variable	Dependent variables							
	Share Responsible Product				Incentivized Start Norm			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Recommend Responsible	0.1850*** (0.0673)				-0.2895*** (0.0873)			
Recommend Harmful		-0.1736*** (0.0556)				0.3450*** (0.1152)		
Difference			0.1644*** (0.0448)				-0.2869*** (0.0707)	
Median Split				0.0877* (0.0474)				-0.1738*** (0.0542)
Constant	0.1360*** (0.0257)	0.2364*** (0.0283)	0.1877*** (0.0258)	0.1629*** (0.0265)	-0.1889*** (0.0400)	-0.3660*** (0.0337)	-0.2675*** (0.0320)	-0.2190*** (0.0353)
No. of observations	4,768	4,768	3,578	4,768	483	483	364	483
No. of markets	69	69	52	69	69	69	52	69
R ²	0.022	0.014	0.042	0.011	0.037	0.039	0.093	0.030

Notes: The variables are measured as the probability (ranging from 0 to 1) that a discussion belongs to each category, based on its overall content. Higher values indicate stronger relevance. **Recommend Responsible:** The discussion recommends trading products without an effect on the donation or considers it inappropriate to trade products with a reduction in the donation. **Recommend Harmful:** The discussion recommends trading products with a reduction on the donation or considers it appropriate to trade products with a reduction in the donation. **Difference:** Recommend Responsible - Recommend Harmful. **Median Split:** Median split of the difference. Robust standard errors are clustered at the market level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.8 Chat content and market behavior across treatments

	Dependent variables					
	Share Responsible Product			Incentivized Start Norm		
	Full Discourse	Seller Only	Buyer Only	Full Discourse	Seller Only	Buyer Only
	(1)	(2)	(3)	(4)	(5)	(6)
Difference	0.1822** (0.0715)	0.1714*** (0.0551)	0.1491 (0.1278)	-0.5309*** (0.1024)	-0.5279*** (0.1390)	0.0165 (0.0970)
Constant	0.2335*** (0.0558)	0.1269*** (0.0237)	0.2000*** (0.0488)	-0.2603*** (0.0492)	-0.1484** (0.0627)	-0.3859*** (0.0499)
No. of observations	1,203	1,198	1,177	119	72	51
No. of markets	17	18	17	17	18	17
R ²	0.048	0.057	0.029	0.318	0.233	0.000

Notes: **Difference:** Recommend Responsible - Recommend Harmful. Robust standard errors are clustered at the market level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C Supplementary Figures and Tables

Figure C.1 Illustration of the market dynamics

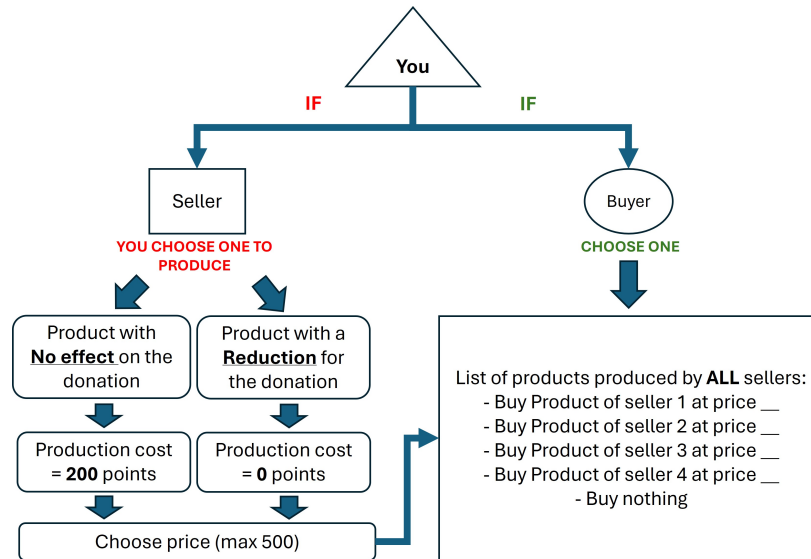
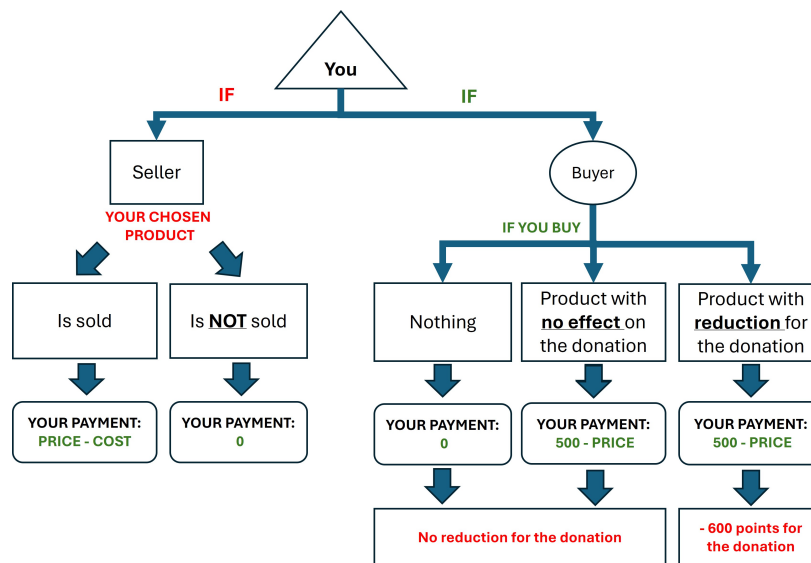


Figure C.2 Illustration of the payments



D Experiment Instructions

D.1 The Market Game

Welcome!

We are pleased to welcome you to this economic study. If you carefully read the following instructions, you can—depending on your decisions and those of the other participants—earn money in addition to the €12 you will receive for participating in this study. Therefore, it is very important that you read the instructions carefully. If you have any questions, please contact us.

Communication with other participants during the study is strictly prohibited. Violation of this rule will result in the participant's exclusion from the study, in addition to the corresponding loss of the final payment.

During the study, we will not speak in €, but in points. All your earnings will be calculated first in points. The points you earn during the study will be converted into € at the end of the study. The following conversion rate will apply: 1000 points = €12.

The main part of the study will consist of 24 periods. At the end of today's study, you will receive in cash the number of points obtained in one of the 24 periods, selected at random, plus the initial endowment of 1000 points (€12) for participating. We will explain the exact procedure of the study on the next pages.

If you have any questions, please raise your hand.

The Study

In today's study, you will participate in a market experiment. In this study, there are two types of participants: sellers and buyers. Participants are divided into markets composed of 7 people: four sellers and three buyers. You will find out whether you are a seller or a buyer on your screen at the beginning of the study. Your role as seller or buyer will not change throughout the study.

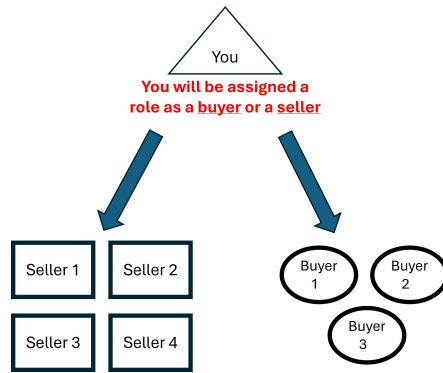


Figure D.1 Role assignment

Donation

After conducting the experiment, for each buyer, we will make a donation of at most 1000 points to GiveDirectly, a charitable organization that allows donors to send money directly to the poorest households in the world. At the end of the instructions, more details about the mission of GiveDirectly are provided. Since there are three buyers in each market, the maximum donation in each market is 3000 points. The exact amount to be donated to GiveDirectly depends on the types of products that sellers and buyers trade. This will be explained in more detail below.

Types of products

The study lasts 24 periods. In each period, each seller chooses a product to sell. Sellers can choose between two types of products:

- “Products with no effect on the donation”
- “Products with a reduction in the donation”

The two products do not differ in anything except in their effect on the donation and in their production cost.

Prices

The price of the products can be any value between 0 and 500.

Production cost

The production cost of a “product with no effect on the donation” is 200 points.

The production cost of a “product with a reduction in the donation” is 0 points.

Product value for buyers

The value of a product for a buyer is always 500 points, regardless of the type of product they buy.

Effect on the donation

- If a buyer purchases a “product with no effect on the donation” or buys no product, the donation will not be reduced and will be 1000 points. For example, the donation will amount to 3000 points if each buyer in a market buys a “product with no effect on the donation.”
- If a buyer purchases a “product with a reduction in the donation,” the donation will be reduced by 600 points, resulting in a donation of 400 points ($1000 - 600 = 400$). For example, the total donation in your market will be 2400 points ($3000 - 600$) if only one buyer decides to buy a “product with a reduction in the donation.”

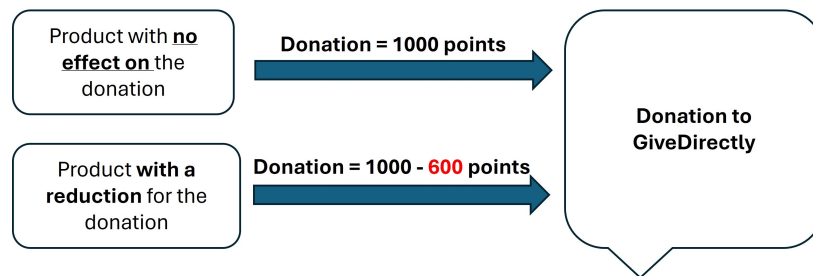


Figure D.2 Effect on the donation

Market activity

The three buyers in a market see the offers of the four sellers in that same market (the price and the type of product) and each of the buyers can accept one offer. The buyers will decide

one after another in a random order. Each buyer can accept only one offer, which means that a maximum of three of the four sellers will be able to sell a product.

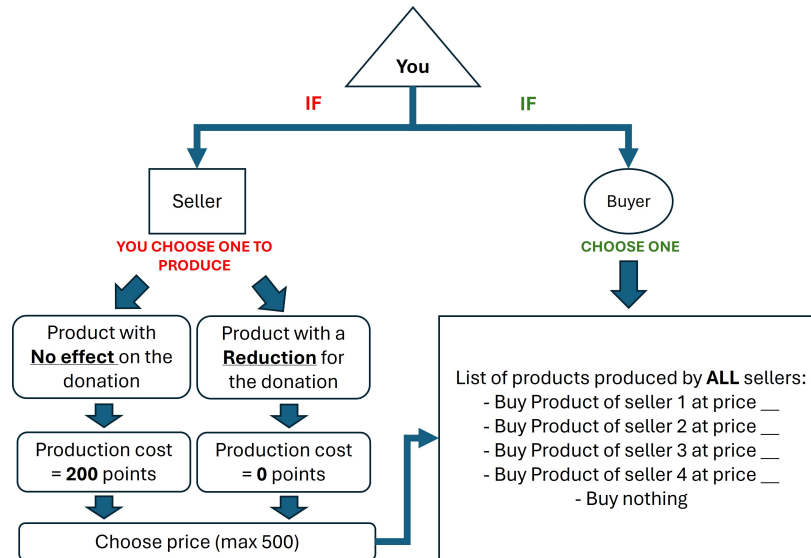


Figure D.3 Illustration of the market dynamics

Earnings

The earnings (in points) of each seller and buyer in a given period are determined as follows:

Seller's earnings

- If a buyer buys their product, the seller's earnings are: product price – production cost.
- If no buyer buys the seller's product, the seller's earnings are: 0.
- Sellers do not incur any production cost if their product is not sold.

Buyer's earnings

- If the buyer buys a product: 500 – product price.
- If the buyer buys nothing: 0 points.

Amount to donate per buyer

- If a buyer chooses to buy a “Product with a reduction in the donation,” $1000 - 600 = 400$ points will be donated.
- If a buyer chooses to buy a “Product with no effect on the donation” or buys nothing, 1000 points will be donated.

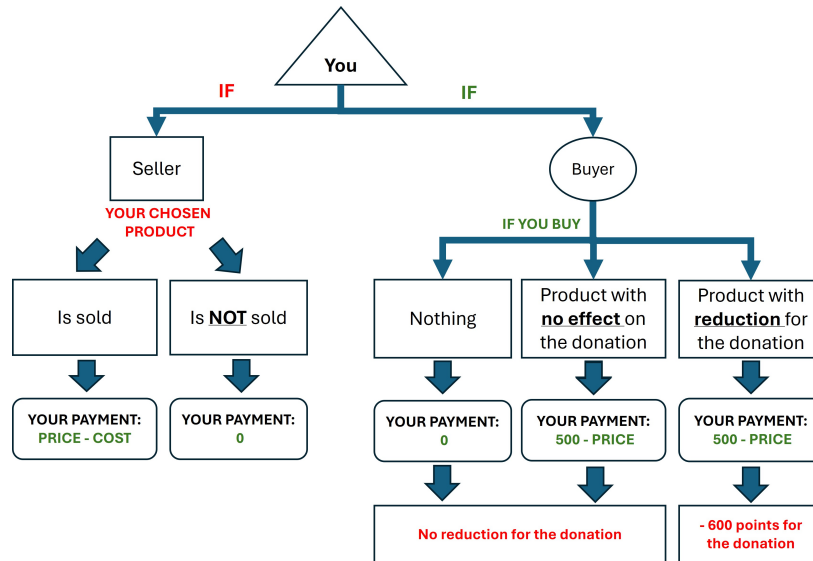


Figure D.4 Illustration of the payments

More about GiveDirectly

GiveDirectly is a non-profit organization that focuses on alleviating poverty through direct cash transfers to people in need. The organization is based on the principle that people in poverty are in the best position to determine their own needs, so giving them money directly is an efficient and effective way to support them. Research cited by GiveDirectly indicates that beneficiaries mainly spend the money on essentials such as food, housing, health care, and education. In addition, many beneficiaries invest in small businesses or other income-generating activities, which can contribute to greater long-term financial stability.

By donating to GiveDirectly, your contribution directly impacts the lives of beneficiaries, allowing them to address their most urgent needs. This approach respects the autonomy of people in poverty and allows a tailored and context-specific use of funds. The potential

benefits include improved living conditions, better access to education and health care, and opportunities for economic advancement, which generate a significant change in the lives of recipients.

Computer procedures

Sellers In each period, sellers indicate which product they want to produce, and then sell, on the following screen:

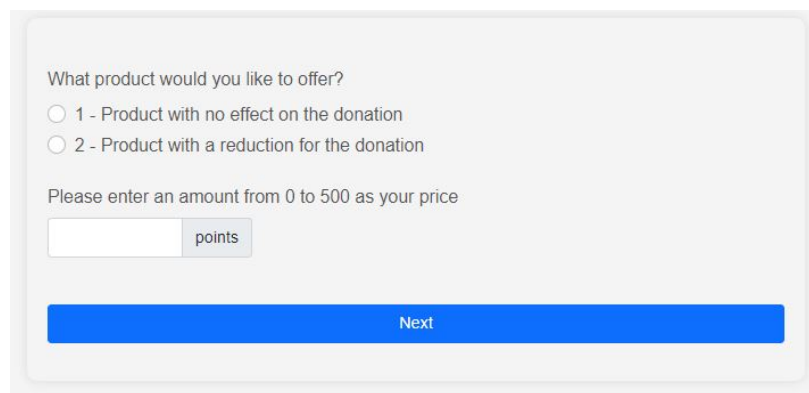
A screenshot of a web-based production window. The window has a light gray background and rounded corners. At the top, it asks "What product would you like to offer?". Below this are two radio button options: "1 - Product with no effect on the donation" and "2 - Product with a reduction for the donation". Underneath the options, it says "Please enter an amount from 0 to 500 as your price". There is a text input field followed by a "points" label. At the bottom of the window is a prominent blue button labeled "Next".

Figure D.5 Production window

Sellers must indicate whether they wish to produce a “product with no effect on the donation” or a “product with a reduction in the donation.” To do so, one must click on the corresponding product type. Sellers must also indicate the price at which they wish to sell the product. The price can range from 0 to 500 points. When producing a “product with no effect on the donation,” sellers must choose a price higher than the production cost (200 points).

Each time a seller has made their decision, they must click the “Confirm” button. Both the product type and the price can be modified until the “Confirm” button is clicked.

When the four sellers have chosen what to produce, they will see a list with the products for sale (the price and the product type) of all the sellers in their market. Here is an example:

The product produced by the seller appears in blue. Sellers can always see on the right side of their screen whether buyers buy their products and in what order they do so.

Products in the Market
The product you produced is shown in blue.

Product Type	Price	Status
Product with a reduction for the donation	???	Purchase order: 3
Product with a reduction for the donation	???	Purchase order: 2
Product with no effect on the donation	???	Purchase order: 1
Product with no effect on the donation	???	Not sold

All the buyers have now made a decision. Press next to continue.

[Next](#)

Figure D.6 Market window

Buyers Buyers see the products for sale available in each period on their screen. The order of the buyers is determined randomly in each period. The first buyer can choose among the four products. The second buyer can only choose among the remaining products, since each product can only be bought by one buyer. The third buyer can only select among the two remaining products. Here is an example:

Each product appears on one line. To buy a product, the buyer must select the corresponding line and then click the “Confirm” button. If a buyer does not want to buy anything, they must click the option BUY NOTHING and press the “Confirm” button.

These are the products in the market. You can now make your decision on what to buy

- Product with no effect on the donation at ??? points price
- Product with a reduction for the donation at ??? points price
- Product with no effect on the donation at ??? points price
- Product with a reduction for the donation at ??? points price
- Buy Nothing

[Next](#)

Figure D.7 Buyer window

When all buyers have made their decisions, all participants will be informed about their earnings and the amount to be donated in that period. After that, the next period will start.

Your payment

The money you earn today will be equal to your earnings from one randomly selected period. This period will also determine the amount of money that we will donate to GiveDirectly. Since it will not be known until the end which period will be used to determine your payment, you must carefully consider your decisions in each of the 24 periods. At the end of the study, the corresponding amount in points will be converted to EUROS and you will be paid in cash.

If you wish to verify our donation to GiveDirectly, at the end of the study you will be able to write your email address, and we will send you the corresponding receipt with the date and the amount donated.

Do you have any questions? If so, please raise your hand. Otherwise, we ask you to click “NEXT.”

Control Questions

P1: Suppose that a seller decides to sell “a product with no effect on the donation” at a price of 400 points and a buyer accepts their offer. What are the seller’s earnings, the buyer’s earnings, and the corresponding donation amount?

- Seller’s earnings: 200; Buyer’s earnings: 100; Donation: 1000
- Seller’s earnings: 350; Buyer’s earnings: 100; Donation: 400
- Seller’s earnings: 150; Buyer’s earnings: 500; Donation: 1000
- Seller’s earnings: 400; Buyer’s earnings: 100; Donation: 400

P2: Suppose that a seller decides to sell “a product with a reduction in the donation” at a price of 400 points and a buyer accepts their offer. What are the seller’s earnings, the buyer’s earnings, and the corresponding donation amount?

- Seller’s earnings: 100; Buyer’s earnings: 100; Donation: 1000
- Seller’s earnings: 400; Buyer’s earnings: 100; Donation: 400
- Seller’s earnings: 200; Buyer’s earnings: 400; Donation: 400
- Seller’s earnings: 200; Buyer’s earnings: 100; Donation: 400

P3: Suppose that a buyer decides not to buy any product. What are that buyer's earnings and the corresponding donation amount?

- Buyer's earnings: 500; Donation: 0
- Buyer's earnings: 0; Donation: 1000
- Buyer's earnings: 0; Donation: 0
- Buyer's earnings: 500; Donation: 1000

P4: Suppose that a seller offers "a product with no effect on the donation" at a price of 400 points and no buyer accepts the offer. What is that seller's payoff?

- Seller's earnings: -200
- Seller's earnings: 0
- Seller's earnings: 500
- Seller's earnings: 400

D.2 Public Discourse

Screen 1

Before beginning the study, some participants will have the opportunity to communicate with each other via a chat for five (5) minutes.

Please use this time to discuss the topic that will be presented on the next page.

The remaining participants must wait until these participants finish chatting.

Click the "Next" button to continue.

Screen 2

You will have five minutes to discuss with other participants how "socially appropriate" or "socially inappropriate" it is to trade a "product with a reduction in the donation." That is, as a buyer or seller, to what extent trading this product aligns with or deviates from what the majority considers "appropriate," "right," or "moral"?

Please write your messages in the chat that will appear on the next page. After typing your message, press the "Enter" key to send it. Each participant has been assigned a random

number that will appear next to their messages. You will see your number when you send your first message. This number is only so that you can identify each other during the discussion. Please do not send messages that could personally identify you.

Screen 3

Chat room

D.3 Norm Elicitation

D.3.1 Before Market Activity

Screen 1

What do you think?

Now we ask you to rate to what extent you consider how “socially appropriate” or “socially inappropriate” it is to trade the “product with a reduction in the donation.” That is, as a buyer or seller, to what extent does trading this product align with or deviate from what you consider “appropriate,” “right,” or “moral”? You can choose among four possible answers: “very socially appropriate,” “somewhat socially appropriate,” “somewhat socially inappropriate,” and “very socially inappropriate.”

I consider that trading the “product with a reduction in the donation” is:

- very socially appropriate
- somewhat socially appropriate
- somewhat socially inappropriate
- very socially inappropriate

Screen 2

What do the others think?

Now we ask you to try to guess which answer was most chosen by the other participants to the previous question. This time we do not ask you to indicate what you personally think is the “correct” answer, but which one you think will be most frequently chosen by the other participants in today’s study.

At the end of the study, we will reveal which answer was selected by the majority of the other participants in this experiment. If you give the same answer as the majority, you will receive an additional 25 points (in addition to your earnings in the main part of the study). Otherwise, you will not receive any additional payment.

For example, suppose you say that the most common answer was “very socially inappropriate.” In that case, you would receive 25 additional points if the most common answer among the other participants is indeed “very socially inappropriate,” but you would receive 0 points if the most common answer is different. Similarly, if you say that the most common answer was, for example, “somewhat socially inappropriate,” you would receive 25 additional points if the most common answer in the study is “somewhat socially inappropriate,” but you would receive 0 points if the most common answer is different.

Screen 3

Now, please indicate your answer. Remember that you will earn additional points if you give the same answer as the majority of the other participants in the study.

Which do you think was the answer most chosen by the other participants in this experiment? Trading the “product with a reduction in the donation” is:

- very socially appropriate
- somewhat socially appropriate
- somewhat socially inappropriate
- very socially inappropriate

Screen 4

You are a Seller/Buyer for the entire duration of the study

D.3.2 After Market Activity

Screen 1

What do you think?

Now we ask you once more to rate to what extent you consider how “socially appropriate” or “socially inappropriate” it is to trade the “product with a reduction in the donation.” That

is, as a buyer or seller, to what extent does trading this product align with or deviate from what you consider “appropriate,” “right,” or “moral”? You can choose among four possible answers: “very socially appropriate,” “somewhat socially appropriate,” “somewhat socially inappropriate,” and “very socially inappropriate.”

I consider that trading the “product with a reduction in the donation” is:

- very socially appropriate
- somewhat socially appropriate
- somewhat socially inappropriate
- very socially inappropriate

Screen 2

What do the others think?

Similarly, we ask you again to try to guess which answer was most chosen by the other participants to the previous question. This time, we do not ask you to indicate what you personally think is the “correct” answer, but which one you think will be most frequently chosen by the other participants in today’s study.

If you select the answer most commonly chosen by the other participants in this experiment, you will earn 25 additional points.

Which do you think was the answer most chosen by the other participants in this experiment? Trading the “product with a reduction in the donation” is:

- very socially appropriate
- somewhat socially appropriate
- somewhat socially inappropriate
- very socially inappropriate

D.4 Dictator Game

Thank-You Gift

Thank you for successfully completing all rounds of the experiment. As a token of our appreciation, you have been awarded an additional 30 points.

You now have the option to keep these 30 points or donate some or all of them to GiveDirectly.

Points to donate (from 0 to 30): _____

D.5 Final Questions

Thank you for participating in today's experiment. Before concluding, we would appreciate it if you would answer the following questions.

Question 1a (Only show to buyers)

As a buyer, how important were the following factors to you when deciding which type of product to buy?

(1 = Not important, 2 = Slightly important, 3 = Somewhat important, 4 = Very important, 5 = Extremely important)

- Your earnings
- The donations
- The sellers' earnings
- The other buyers' earnings

Question 1b (Only show to buyers)

As a buyer, how important were the following factors to you when deciding *not* to buy anything? (Answer only if you ever decided not to buy anything)

(1 = Not important, 2 = Slightly important, 3 = Somewhat important, 4 = Very important, 5 = Extremely important)

- Your earnings
- The donations
- Only products with no effect on the donation were available
- Only products with a reduction in the donation were available

Question 1c (Only show to sellers)

As a seller, how important were the following factors to you when deciding which type of product to produce?

(1 = Not important, 2 = Slightly important, 3 = Somewhat important, 4 = Very important, 5 = Extremely important)

- Your earnings
- The donations
- Probability of selling your product
- Producing the same product as the other sellers
- The order in which your product would be purchased
- Not being able to sell your product
- The product price

Question 2

If there were other factors that were relevant, please comment briefly here:

Question 3

How willing are you to perform good deeds without expecting anything in return?

(0 = completely unwilling, 10 = very willing)

Score:

0 1 2 3 4 5 6 7 8 9 10

E Pre-Analysis Plan

E.A Experiment Design

E.A.1 Description of the experiment

Our experiment involves a product market game in which participants, assigned to the roles of either sellers or buyers, trade two types of products: a low-cost product that generates a negative externality by reducing a donation fund and a high-cost product that is socially responsible and imposes no harm on the donation. Sellers act first, deciding whether to produce the low-cost, harmful product or the high-cost, socially responsible product, and then setting a price. Buyers then decide which product to purchase or whether to abstain from buying altogether. Each market comprises seven participants—four sellers and three buyers.

Before the market game begins, participants will read the instructions. These instructions include information about their potential roles as sellers or buyers, how decisions in the market game are made, and how payments are determined. Participants will also complete a series of control questions to ensure they fully understand the experimental procedures.

Additionally, prior to the start of the market game and before their roles as sellers or buyers are revealed, some participants within each market will engage in a chat discussion about the socially responsible course of action in the game. During this discussion, they will deliberate on the social appropriateness of trading the harmful product. Afterward, participants will individually rate how “socially appropriate” or “socially inappropriate” they perceive it is to trade the harmful product. In an incentive-compatible manner, participants will also predict how they believe others in the experiment answered the same question. Once these steps are completed, participants are informed of their roles as sellers or buyers, and the market game begins.

The game consists of 24 rounds. At the end of each round, participants are informed about their individual earnings for that round and the total amount of money donated to the fund. To determine final payments, one round is randomly selected at the end of the experiment, and the outcomes from that round are used to calculate participants’ total earnings and the donation.

After completing the 24 rounds, participants are asked once more to rate how socially appropriate they consider trading the harmful product. They are also asked once again to predict how they believe other participants responded to the same question.

The experiment concludes with a dictator game, where participants have the opportunity to donate additional money to the donation fund. Following this, they answer questions about their motivations for producing or purchasing socially responsible or harmful products. Finally, demographic information is collected, and participants are informed of their total earnings from the experiment, including the amounts donated.

E.A.2 Treatment assignment

In our experiment, participants are randomly assigned to one of four treatments. By using a 2x2 between-subjects design, we vary whether (a) sellers and (b) buyers can chat with other market participants prior to the first round. This results in four treatment conditions:

- (i) *No Discourse*: neither side of the market can chat.
- (ii) *Full Discourse*: both sides of the market can chat with everyone else.
- (iii) *Seller-Only Discourse*: sellers can chat with sellers.
- (iv) *Buyer-Only Discourse*: buyers can chat with buyers.

This design enables us to explore how different forms of discourse influence market outcomes and individual behavior.

For our analysis, we will consider all subject responses as valid, unless otherwise specified in the following sections.

E.B Main Outcome

The main outcome variable will be a binary indicator that equals 1 if a socially responsible product is traded and 0 if a harmful product is traded. We exclude observations where no purchase is made.

E.C Main Explanatory Variable

The treatment conditions: we will compare the main outcome across the different treatments.

E.D Main Analysis

1. Public Discourse: does public discourse impact socially responsible behavior?

- (a) We will report the proportion of socially responsible products traded (our main outcome) in each market by treatment.
- (b) To assess the impact of the different treatments on the proportion of socially responsible products traded, we will use an OLS regression:
 - i) We will regress the outcome variable on treatment indicators. Specifically, the model will include a Full Discourse treatment dummy variable, which equals 1 if the traded product corresponds to the Full Discourse treatment condition, and 0 otherwise; a Seller-Only Discourse dummy variable, which equals 1 if the traded product corresponds to the Seller-Only Discourse treatment condition, and 0 otherwise; and a Buyer-Only Discourse dummy variable, which equals 1 if the traded product corresponds to the Buyer-Only Discourse treatment condition, and 0 otherwise.
 - ii) The No-Discourse treatment will serve as the reference group in the regression.
 - iii) Standard errors will be clustered at the market level.
- (c) Additionally, we will test whether the Buyer-Only and Seller-Only Discourse conditions differ from the Full Public Discourse condition on the main outcome. To do so we will use an OLS regression:
 - i) We will regress the outcome variable on the treatment indicators. The model will include a No Discourse treatment dummy variable, which equals 1 if the traded product corresponds to the No Discourse treatment condition, and 0 otherwise, a Seller-Only Discourse treatment dummy variable and a Buyer-Only Discourse treatment dummy variable.

- ii) In this case, the Full Discourse treatment will serve as the reference group in the regression.
 - iii) Standard errors will be clustered at the market level.
- (d) Finally, we will test whether the main outcome differs significantly between the Seller-Only Discourse condition and the Buyer-Only Discourse condition. To do so we will use an OLS regression:
- i) As in point (b) and (c), we will regress the outcome variable on the treatment indicators. The model will include a No Discourse treatment dummy variable, a Full Discourse treatment dummy variable, and a Buyer-Only Discourse dummy variable.
 - ii) In this case, the Seller-Only Discourse treatment will serve as the reference group in the regression.
 - iii) Standard errors will be clustered at the market level.
- (e) If we are unable to detect that the Buyer-Only and Seller-Only Discourse conditions differ from the No-Discourse condition (from tests in point (b) and (d)) we will combine the Buyer-Only and Seller-Only treatments to increase statistical power. However, we will only consider these additional tests when:
- i) Buyer-Only = No-Discourse and Seller-Only = No-Discourse
 - ii) Buyer-Only > No-Discourse and Seller-Only = No-Discourse but Seller-Only = Buyer-Only
 - iii) Buyer-Only = No-Discourse and Seller-Only > No-Discourse but Seller-Only = Buyer-Only

To do so:

- i) We will create a variable named Partial Discourse treatment dummy that equals 1 if the traded product corresponds to either the Seller-Only or Buyer-Only treatment condition, and 0 otherwise.

- ii) As in point (b), we will regress the outcome variable on the treatment indicators. The model will include a Full Discourse treatment dummy variable and a Partial Discourse treatment dummy variable, where the No Discourse treatment will serve as the reference group in the regression. We also compare Partial Discourse to Full Discourse by changing the reference group to Full Discourse, similar to point (c) above.
 - iii) Standard errors will be clustered at the market level.
- (f) We will use one-sided p-values to test the following directional hypotheses in the regressions outlined above:
- i) H.D.1: Full Discourse $>$ No Discourse
 - ii) H.D.2: Full Discourse $>$ Seller-Only Discourse
 - iii) H.D.3: Full Discourse $>$ Buyer-Only Discourse
 - iv) H.D.4: Seller-Only Discourse $>$ No Discourse
 - v) H.D.5: Buyer-Only Discourse $>$ No Discourse

We will use a two-sided p-value to test the following hypothesis, as we do not have a clear expectation regarding the direction of the effect:

- vi) H.D.6: Seller-Only Discourse \neq Buyer-Only Discourse

In the case of conducting the tests from point (e), we will use one sided p-values:

- vii) H.D.7: Partial Discourse $>$ No Discourse
- viii) H.D.8: Full Discourse $>$ Partial Discourse

E.E Secondary Analysis

1. Public discourse dynamics: does the public discourse effect vary over time?

- (a) We will report the proportion of socially responsible products traded in each market (our main outcome) over time, separately for each treatment.

- (b) To assess the dynamic effects of the different treatments on the proportion of socially responsible products traded, we will use an OLS regression.
 - i) We first divide the sample into two-time blocks, with each block consisting of 12 rounds: Block 1 (Rounds 1–12), Block 2 (Rounds 13–24).
 - ii) We will regress the main outcome variable on the treatment dummies, block dummies and the interaction between the treatment dummies and the block dummies.
 - iii) The reference groups for the regression will be the No-Discourse treatment condition and Block 1.
 - iv) Standard errors will be clustered at the market level.
- (c) Similar to the main analysis (D.1), we will adapt points D.1 (c) and D.1 (d) to incorporate dynamic effects based on E.1 (b).
- (d) Following instructions from point D.1 (e), if the right conditions are met, we will repeat tests from points E.1 (b) to E.1 (c) – but excluding analysis D.1 (d) mentioned in E.1 (c) – to explore dynamic effects in the partial treatment.
- (e) We will use two-sided p-values for all coefficient estimates in these regressions.
- (f) We will allow ourselves to redefine the grouping of rounds into Block 1 and Block 2 (as described in E.1 (b)) by changing the cutoff point that determines which rounds belong in which block, in response to visual inspection of the data. This means the number of rounds included in Block 1 and Block 2 may differ based on the specific context or analysis needs. This gives us the flexibility to consider different types of dynamics.

2. Seller Behavior

- (a) We will replicate the analyses outlined in sections D.1 and E.1 from the perspective of sellers. For this purpose, we will redefine the main outcome variable as a binary variable, which will take the value 1 if a seller offers a socially responsible product and 0 otherwise. We expect very similar results for our main analysis (D.1) but possibly different dynamics (E.1).

3. Prices and Profits: are buyers willing to pay higher prices to trade the socially responsible products? Are sellers willing to forgo profits by opting to trade socially responsible products?

- (a) We will report the average prices and profits separately for each treatment and product type.
- (b) We will investigate whether buyers are willing to pay higher prices and whether sellers are willing to forgo profits by opting to trade socially responsible products. To do so we will use an OLS panel regression model with participant level fixed effects and round fixed effects
 - a. To evaluate buyers' willingness to pay higher prices, we will regress the main outcome variable on two variables: (i) a variable indicating the lowest price of a harmful product available for a given buyer in a given round and (ii) a variable indicating the lowest price of a socially responsible product available for a given buyer in a given round.
 - i) We create variables (i) and (ii) as follows:
 1. First, for each seller in a market we define the type of product they created in a given round.
 2. Second, for each buyer in a market in a given round, we define the type of product available from each of the sellers (excluding products already bought by other buyers). That way, we can determine the type of product from a given seller in a given round that was available for a given buyer. This allows us to see the type of products that the buyer could choose from (and from which seller it came from).
 3. Third, we create a new variable indicating the price of each of the available products.
 4. Fourth, we create both variables by determining the cheapest product available for a given buyer in a given round, separately for each product type.

- ii) The model omits cases in which a buyer did not purchase a product and cases in which either only the responsible or only the harmful products were available to a buyer.
 - b. To evaluate sellers' willingness to forgo profits, we will regress the outcome variable, a variable taking on value 1 if a seller offered a responsible product and 0 otherwise, on a variable that indicates the average realized profit difference between offering a socially responsible product and offering a harmful product in the preceding period.
 - i) We create that variable in the following way:
 1. First, we calculate the profit obtained by each seller in a given market in a given round.
 2. Second, we calculate the average profit of sellers in a given market in a given round from producing the socially responsible product.
 3. Third, we repeat step 3 but for the harmful product.
 4. Fourth, we generate our variable of interest by subtracting step 2 from step 3 in the prior round.
 - ii) If an offer is not accepted, the seller's profit equals zero.
 - iii) The model omits the first round due to not having a prior round. It also omits cases in which, in the previous round, only one of the two product types was offered.
 - c. Standard errors will be clustered at the market level.
- (c) We will repeat the analysis in point E.3. (b) separately by treatment group
- (d) We will use one-sided p-values to test the following hypotheses in the regressions outlined above
- a. H.E.3.1: Lowest price of the responsible product < 0
 - b. H.E.3.2: Lowest price of the harmful product > 0
 - c. H.E.3.3: Expected responsible product profit premium > 0

4. Social Norms: does public discourse impact perceived social appropriateness of trading harmful products?

- (a) We will report the average appropriateness rating of trading harmful products under the different treatment conditions “before” and “after” the market interaction. We refer to these as the “prior” and “posterior” ratings.
 - a. We use the incentivized measure of the average appropriateness rating of trading harmful product as the outcome variable.
- (b) We measure the average appropriateness rating of trading harmful products following Krupa and Weber (2013).
- (c) We will examine whether the average appropriateness rating of trading harmful products is significantly lower in the Full Discourse, Seller-Only and Buyer-Only treatment conditions compared to the No-Discourse treatment condition. To do that, we will use OLS regressions:
 - a. First, we will regress the “prior” rating as the outcome variable on the treatment dummies.
 - b. Second, we will regress the “posterior” rating as the outcome variable on the treatment dummies.
 - c. Third, we will regress “both” ratings as the outcome variable on the treatment dummies, a posterior variable (0: prior rating, 1: posterior rating) and the interaction between the treatment dummies and the posterior variable.
 - d. The reference groups for these regressions will be the No-Discourse treatment condition and the prior rating (for the third regression).
 - e. Standard errors will be clustered at the market level.
- (d) Additionally, we will test whether, in the Seller-Only and Buyer-Only treatment conditions, the average ratings of buyers and sellers converge after the market interaction. To do that, we will use OLS regressions:
 - a. We will perform the same regressions stated in point E.4. (c) but separately for buyers and sellers.

- b. Standard errors will be clustered at the market level.
- (e) We vary the reference group in E.4 (c) and E.4 (d) to compare the ratings with the other treatment groups (similar to D.1 (c) and (d)).
- (f) We will use either one-sided or two-sided p-values to test the following hypotheses in the regressions outlined above
 - a. For regressions a and b in point E.4 (c) E.4 (d) and E.4.(e) we perform the same hypotheses as in the main analysis (D.1 (f)) but with opposing sign.
 - b. For the regression involving the pooled sample with both buyers and sellers and both ratings (regression c in point E.4 (c) including changes in the reference treatment specified in E.4.(e)) we will do two-sided tests on all coefficients.
 - c. For the regressions that involve only buyers or only sellers and both ratings (regression c in point E.3 (d) including changes in the reference treatment specified in E.3 (e)) we will test:

- **No Discourse and Prior as reference group:**

For both the sellers and buyers regression:

- i) H.E.4.1: Full Discourse x Posterior $\neq 0$ (two-sided)

For the sellers regression:

- ii) H.E.4.2: Seller-Only Discourse x Posterior > 0 (one-sided) – Convergence with respect to the prior
- iii) H.E.4.3: Buyer-Only Discourse x Posterior < 0 (one-sided) – Divergence with respect to the prior

For the buyers regressions:

- iv) H.E.4.4: Seller-Only Discourse x Posterior < 0 (one-sided) - Divergence with respect to the prior
- v) H.E.4.5: Buyer-Only Discourse x Posterior > 0 (one-sided) - Convergence with respect to the prior

- **Full Discourse and Prior as reference group:** same as previous point but opposite sign

- **Seller-Only Discourse and Prior as reference groups:**

For the sellers regression:

- vi) H.E.4.6: Buyer-Only Discourse x Posterior < 0 (one -sided) – Convergence with respect to the prior

For the buyers regression:

- vii) H.E.4.7: Buyer-Only Discourse x Posterior > 0 (two-sided) – Convergence with respect to the prior

- d. We will also perform a two-sided t-test to compare if the posterior rating of the seller is different from the posterior rating of the buyer under the Seller-Only Discourse treatment condition. We repeat the same exercise for the Buyer-Only Discourse treatment condition.

5. Dictator Game: does socially responsible behavior increase donations?

- (a) We will report the fraction of money donated to a charity in a post-experimental dictator game under the four different conditions mentioned above.
- (b) We will test whether being exposed to public discourse generates spillover effects and increases donations. To test this, we will employ an OLS regression.
 - i) We will regress the fraction of money donated in the dictator game by a participant as the outcome variable on the treatment dummies.
 - ii) The reference groups for these regressions will be the No-Discourse treatment condition.
 - iii) Standard errors will be clustered at the market level.
- (c) We will also repeat this analysis by changing the reference treatment in line with our main analysis (D.1 (c) and D.1 (d)).
- (d) The corresponding hypotheses are the same as the ones in the main analysis (D.1 (f))

6. Self-reported reasoning: are self-reported reasons in the questionnaire correlated with socially responsible behavior?

- (a) We will test whether self-reported reasoning correlates with socially responsible behavior. Specifically, we will analyze whether participants' justifications for their decisions correlate with their choice to trade socially responsible products, providing insights into the motivations driving their behavior.
 - i) We will use OLS regressions. We will regress the main outcome variable on their responses to the questions asked at the end of the experiment.
 - ii) We will also regress the fraction of times a buyer decided not to buy any product as the dependent variable on their responses to these questions.
 - iii) Standard errors will be clustered at the market level.
 - iv) We will perform these analyses for the full sample as well as separately by treatment.
- (b) We will use two-sided p-values.

7. Content analysis: do the topics discussed in the chat predict the decisions of buyers and sellers?

- (a) We plan to use AI to analyze the content of the chats and assess whether it can predict market behavior. The specific methodology for this analysis is yet to be determined.

E.G Robustness check

We will repeat the main analysis (section D) with probit and GLS (random effects) regressions as robustness tests (instead of using OLS).

E.H Power Calculations

We aim to recruit 504 participants, with 126 participants per treatment group, distributed across 18 markets per treatment. This sample size is designed to achieve a statistical power

of 0.80 to detect a treatment effect of 17 pp (percentage points) in the fraction of socially responsible products traded.

Our power calculations are based on Bartling et al. (2024), specifically Study 2, which is most comparable to our experimental setup. In their study, Bartling et al. (2024) also compare a public discourse to a no-discourse condition in a very similar market experiment. They observed a 33 pp increase in the fraction of socially responsible products traded (from 50% to 83%) under the public-discourse condition compared to the no-discourse condition. We chose Study 2 as the basis for our analysis because it has the most similar design and because it presents a more conservative estimate of treatment effects compared to Study 1. Using the replication data provided by Bartling et al. (2024), we calculate the power needed to detect different treatment-effect sizes under an OLS framework.

We calculate power as follows: first, we randomly draw a sample of buyers from Study 2 (public-discourse and no-discourse treatments) with replacement, ensuring a balanced allocation of buyers across treatments. These buyers are then assigned to different markets of equal size (three buyers each). For each of these samples, we then run the regression analysis specified in Section D.1(b) above. Power is defined as the proportion of simulated samples in which the public-discourse treatment is significantly higher on the main outcome (socially responsible behaviour) than the no-discourse treatment ($p < 0.05$, one-sided test). We perform 1,000 simulations and do so for various different sample sizes. This procedure allows us to calculate power for a treatment effect of the same size as Bartling et al. (2024).

To explore power for smaller effect sizes, we also conduct simulations in which we adjust behaviour in the public-discourse treatment to generate smaller effects. In particular, in our simulations we superimpose “harmful” product choices on a fraction of buyer-round observations. Power is then recalculated using the same method described above.

The power calculations indicate that our design will have sufficient power to detect treatment effects of 17 pp (relative to baseline) with a power of 0.80 for our intended sample size of 18 markets per treatment (see Table E.1). This allows us to reliably detect effects similar to Bartling et al. (2024, 33 pp) but also detect the smaller effects we would expect for the buyer-only discourse and seller-only discourse conditions, up to approximately half the size of the original effect (17 pp).

Table E.1 Power calculations for the Public-Discourse vs. No-Discourse condition

Effect size	$N = 16$ markets	$N = 17$ markets	$N = 18$ markets
33 pp	0.996	0.997	0.997
17 pp	0.770	0.782	0.814
15 pp	0.698	0.699	0.719
13 pp	0.584	0.621	0.640

Notes: The table presents results from three sets of simulations that vary the sample size from 16 to 18 markets for both public-discourse and no-discourse conditions. The outcome variable is a dummy equal to 1 if a buyer purchases a socially responsible product and 0 if a harmful product is traded. Each calculation is based on 1,000 simulated samples. The effect sizes are the differences between public-discourse and no-discourse averages across all simulation samples. Power is the fraction of simulations in which an OLS regression of the outcome on the public-discourse indicator produced a significant result at the 5% level (one-sided test). Standard errors are clustered at the market level.