UNIVERSITÀ DEGLI STUDI DI VERONA

DEPARTMENT OF ECONOMICS

DOCTORAL PROGRAM IN

Economics and Finance

Cycle: XXXVII

Human Behavior in Competitive and Informational Environments: Evidence from Experimental and Empirical Analyses

S.S.D. ECON-01/A

Coordinator: Prof.ssa Athena Picarelli

Signature ____________

Tutor: Prof. Marco Piovesan

Signature ____________

Tutor: Prof. Simone Quercia

Signature ____________

Doctoral Student: Dott. Alessandro Barbazeni

Signature ____________

UNIVERSITÀ DEGLI STUDI DI VERONA

DEPARTMENT OF ECONOMICS

DOCTORAL PROGRAM IN

Economics and Finance

Cycle: XXXVII

Human Behavior in Competitive and Informational Environments: Evidence from Experimental and Empirical Analyses

S.S.D. ECON-01/A

Alessandro Barbazeni

Contents

List of Figures

List of Tables

Preface

This thesis uses experimental and empirical analyses to uncover behavioral insights about competitive and informational environments. The structure of this work consists of two independent parts that together capture distinct aspects of human decision-making. The first part, comprising Chapters 1 and 2, examines the behavioral spillovers of tournament incentives on prosocial behavior, experimentally testing how exposure to competitive settings influences individuals' willingness to cooperate. The second part, presented in Chapter 3, establishes the groundwork for understanding emotional responses to misinformation, analyzing how the emotional content of information correlates with its accuracy.

Understanding human behavior in social and economic contexts requires studying the trade-off between self-interest and collective benefits. In economics, incentives are key drivers of individual decision-making, but when multiple forms of monetary rewards coexist, behavior can become unpredictable, particularly under conditions of uncertainty. Cooperation offers a compelling example. Originally studied from an evolutionary perspective to understand some of the earliest life forms, the notion of cooperation evolved in human history to describe more complex social behaviors, such as collective hunting and organized warfare. As societies developed, the study of cooperation has laid the foundation for achieving collective efficiency, measured by incentive mechanisms that set the individual cost of cooperating higher than the personal benefit, yet lower than the overall benefit to the society. Nevertheless, contemporary economies reward individual achievements, thereby interfering with the very mechanisms that sustain cooperation.

The existing literature highlights that the benefits of tournament incentives may come at the cost of undermining prosocial behavior. Exposure to competitive environments can crowd out cooperative norms, reducing individuals' willingness to contribute to shared projects. Yet, more recent evidence questioned the existence of such spillover effects, showing that not all competitive settings lead to the erosion of cooperation. Two central questions arise from this debate: (i) Why do tournament incentives generate negative spillovers on cooperation decisions in some settings and not in others? (ii) Furthermore, because both competition and cooperation incentives yield desirable outcomes, can they be jointly implemented without efficiency losses? In the first two chapters of this dissertation, I address these questions.

Chapter 1 examines whether variations in competitive environments explain differences in the magnitude of spillover effects. Specifically, it distinguishes between standard and unfair competitions to examine whether previously observed effects stem from the hostility of the competitive context. Using a controlled laboratory experiment based on an all-pay auction competition with and without sabotage activities, the study measures how exposure to destructive competition influences subsequent cooperation in a public goods game. The results show that competition generally reduces cooperation, consistent with findings supporting this hypothesis, but that the possibility of sabotage does not further exacerbate spillovers.

Chapter 2 builds on the assumption that incentives to compete and cooperate produce "Mixed Signals" (Gneezy, 2023). In his book, Prof. Gneezy shows how incentive systems should avoid embedding opposing goals: for example, rewarding individual achievement while emphasizing collective success may create conflicting motivations. Yet, many organizational settings simultaneously promote internal competition (e.g., promotion tournaments) and teamwork to benefit from the efficiency improvements stemming from both incentive systems. This chapter experimentally tests whether embedding prosocial signals within a competitive environment, such as donation programs, can offset the negative spillovers of competition on cooperation. I replicate experimental conditions documenting spillovers from competition on cooperation, introducing variations where competition winners are given the opportunity to donate a share of their earnings to charity, generating a visible prosocial signal observed by competition losers. The results provide a well-powered replication of previous findings but fail to reproduce the negative spillover effect of competition on cooperation. Furthermore, donation opportunities do not significantly alter cooperation levels among participants. While winners' donation choices correlate with their own prosocial behavior, these signals do not appear to influence others' cooperative behavior. These findings suggest that integrating prosocial elements into competitive environments may not automatically restore or reinforce cooperative norms.

Altogether, Chapters 1 and 2 highlight the need for further research into the mechanisms underlying the interplay between competition and cooperation, with implications for designing incentive structures that enable the simultaneous implementation of both tournament and cooperative rewards.

Finally, Chapter 3 shifts focus from economic competition to the domain of information sharing, where emotions and behavioral biases drive the spread of misinformation. Digital platforms have transformed how information is produced and consumed, enabling false or misleading content to spread and shape individuals' and collective beliefs. Furthermore, the proliferation of AI-generated content has further amplified these dynamics, expanding opportunities for the creation and dissemination of misinformation. While existing research has documented that false news spreads faster and deeper than true news, the mechanisms underlying this phenomenon remain poorly understood. One critical and under-explored factor is the emotional composition of misinformation. This chapter investigates how news' emotional content correlates with the factual accuracy of information, as well as how these relationships vary across political contexts. Using a large dataset of fact-checked headlines, the study applies a novel emotion classification method to measure the emotional tone of textual content. The results reveal that high emotional intensity and negativity are strongly associated with misinformation, with anger, surprise, fear, and disgust being the most predictive emotions. The analysis also uncovers partisan asymmetries, informing future experimental research on the behavioral mechanisms that drive misinformation. Ultimately, the chapter highlights the importance of identifying the mechanisms that drive the spread of misinformation by analyzing both the consumers and producers of information, focusing on factors such as attention on the consumer side and the constraints faced by producers of accurate information.

Together, the chapters in this thesis offer evidence on how competition, cooperation, and emotions drive decision-making. The results underscore the complexity of behavioral responses in economic and informational environments, and highlight the need to test new incentive schemes while further investigating the mechanisms that drive the production and spread of misinformation.

Chapter 1

Destructive Competition and Its Effects on Cooperation

A. Barbazeni

Abstract

We test experimentally the impact of reducing competitors' chances of winning (sabotage) on cooperation. Participants compete with or without the sabotage option and play a public good game before and after the exposure to a competitive environment. Contributions to the public good decrease by 9% of the initial endowment when observing pre-post cooperation decisions. Nonetheless, we find no evidence that competition with sabotage further reduces contributions to the public good. Yet, the amount of sabotage received negatively predicts contribution within treated participants.

Keywords: Competition; Cooperation; Sabotage; All-Pay Auction.
JEL Codes: C92, D44, H41.

1.1 Introduction

In many economic and organizational settings, individuals' choices are driven by both competitive and cooperative incentives. Competitive environments can incentivize individual effort and productivity, while cooperation among agents sharing a common project can promote efficient outcomes. However, previous studies have detected negative spillover effects of competition on cooperation (Brandts and Riedl, 2020; Buser and Dreber, 2016), highlighting how tournament incentives may undermine prosocial behavior.

More recently, a meta-analysis on the effect of competition on moral behavior found only a small adverse effect of competition, attributing the broader inconclusiveness of the literature to substantial design heterogeneity across experimental protocols (Huber et al., 2023). Yet, Huber et al. (2023) do not identify which specific features of competitive environments drive this heterogeneity, as their crowd-sourced design intentionally leaves all design choices to independent research teams. In this paper, we investigate one such feature: the intrinsic fairness of the competitive environment. Specifically, we compare two competitive settings that differ in their degree of fairness: a "standard" (fair) competition and a "hostile" (unfair) competition in which participants can actively undermine opponents' performance, thereby reducing their chances of success (sabotage).

Understanding these mechanisms requires data on behaviors that are rarely observable in real-world settings. Individual and team performances can be monitored by directly observing behavior in the field. However, sabotage activities are most of the time unobserved. Agents practice sabotage secretly to avoid the repercussions of their unethical behavior. For this reason, we study the impact of competition with sabotage on cooperation in a laboratory experiment, where we can observe individuals' effort, sabotage, and cooperation decisions ceteris paribus. We randomly allocate students to engage in a multi-period All-Pay Auction (APA) contest (Riley and Samuelson, 1981), either with or without sabotage activities, and measure their willingness to cooperate in a public goods game (PGG) both before and after the exposure to the competitive environment. We can isolate the psychological spillover effect of sabotage on cooperation by re-matching participants in groups of perfect strangers in the second cooperation decision. In this way, we rule out mechanisms such as direct reciprocity, reputation, and end-game effects and measure the impact of previous experiences in a destructive competitive environment on cooperation. Furthermore, we can measure the effect of tournament incentives on cooperation by analyzing cooperation decisions before and after the tournament in a within-subject analysis.

We find that participants in both competition with and without the sabotage option contribute significantly less in the second PGG compared to the one played before the exposure to the competitive environment. However, when sabotage is embedded into the competition, the two groups do not differ in their contribution levels. Finally, in a non-preregistered analysis, we find that among participants exposed to the sabotage competition, individuals who receive more sabotage contribute significantly less.

Sabotage in competition is typically detrimental, as agents can increase their probability of success by interfering with their peers' performances, withholding and falsifying information, or tampering with others' activities. Instances of its undesirable effects relate to the performance of organizations (patent race, market shares), workers (promotions, bonuses), and students (sports promotions, academic success). One popular example concerns Uber employees accused of ordering and canceling rides on Lyft to reduce drivers' availability and frustrate customers. To mitigate such behaviors, companies and governments often invest substantial resources in legal and practical solutions, increasing their overall costs. In addition, if incentives to compete are sufficiently high and agents compete over extended horizons, then sabotage has the potential to hinder the development of cooperative attitudes. It is therefore important to understand to what extent the detrimental effects of sabotage in competition also threaten efficiency from cooperation.

We contribute to the literature by analyzing the efficiency implications of fair versus unfair tournament incentives, with a focus on prosocial outcomes. Studies show that the adoption of tournament incentives leads agents to compete more aggressively. When feasible, agents exploit sabotage activities to impair competitors' performance (Charness et al., 2014; Chen, 2003; Harbring and Irlenbusch, 2008, 2011; Konrad, 2000), supporting theories for which equal pay leads to a more efficient outcome (Lazear, 1989). Another feature of a competitive environment is that agents generally prefer competing with weaker opponents and sabotage them for their personal gain, although a notable minority abstain from sabotaging and show some interest in fair outcomes (Buser et al., 2021). Additionally, a non-zero share of agents are willing to engage in competitive settings where sabotage activities are possible (Buser and Sangi, 2025), suggesting a positive selection into destructive competitive environments, which also depends on the amount of information on the competitors (Gürtler et al., 2013).

Concurrently, past research found that tournament incentives carry negative spillovers on prosocial behavior when cooperation decisions are taken after exposure to a competitive environment. Buser and Dreber (2016) find that experiencing competition substantially reduces the social surplus of cooperation. In their experiment, participants first compete in a real effort task and play a standard public good game after being reassigned to new groups. More recently, Li et al. (2023) identified a negative spillover effect of competition on cooperation among participants who made their cooperation decision with a previous competitor. They find that both the experience of defeat and uncertainty in tournaments reduce contributions in a public good game played with the same recent opponents. In addition, Lien et al. (2021) examine how different types of contests (all-pay auction contests, Tullock contests, and proportional prize contests) influence decisions in social dilemma situations. They find supportive evidence that competition leads agents to be less prosocial. Other studies find that competition affects prosocial behaviors in more specific, yet related, contexts. For instance, individuals who have engaged in competitive interactions within a market setting tend to decrease their contributions when later participating in a two-person linear public goods game with previous competitors, but not when cooperation decisions are taken after forming new groups (Brandts and Riedl, 2020; Cason and Gangadharan, 2013). Supporting this evidence, Moyal and Ritov (2020) and Savikhin and Sheremeta (2013) found that participation in contests also reduces people's willingness to donate. Differently, other studies conclude that when the competitive and cooperative games are played simultaneously and share a common endowment, then competition lowers overbidding, but has no impact on cooperation decisions (Godoy et al., 2013; Savikhin and Sheremeta, 2013).

By contrast, competition can be efficient when it operates within a cooperative game. Markussen et al. (2014) demonstrate that inter-group competition enhances cooperation by incentivizing relative group performance. They also show that competition is preferred to other institutions when subjects vote on which set of rules the game is played. Similarly, Augenblick and Cunha (2015) reveal that contributions to a political party escalate when agents are informed of the contributions made to the opposing party. Finally, Kosfeld and Von Siemens (2011) develop a theoretical model suggesting that market competition enhances workers' cooperation in firms where the personnel selection process emphasizes teamwork skills, driving the success of companies such as Southwest Airlines, which has become one of the most efficient airlines in the U.S. market. Past studies also document that within-group competition leads to an increase in cooperation. When individuals compete for a better return in a public good game, agents increase their contributions (Angelovski et al., 2019; Bergantino et al., 2021; Colasante et al., 2019, 2017). However, when the risk associated with the cooperation decision is excessively high (for instance, heterogeneous and uncertain returns), free-riding is observed with greater frequency (Bergantino et al., 2021; Canegallo et al., 2008; Colasante et al., 2017).

Finally, a recent meta-analysis (Huber et al., 2023) finds that competition has a small but negative effect on moral and prosocial behavior, with substantial heterogeneity across experimental designs. This heterogeneity partly reflects differences in the experimental designs used to study the relationship. In many studies, participants are first exposed to a competitive manipulation and subsequently engage in an economic game to assess its effect on moral behavior. Some designs use prosocial outcomes, such as the trust game or dictator game, as proxies for moral behavior[1]. The results, however, do not reveal a consistent pattern, collectively suggesting only a modest effect of competitive settings on prosocial outcomes, as further documented in the replication exercise reported in Chapter 2.

Building on this framework, we examine how the nature of competition (fair versus unfair) shapes subsequent cooperative behavior. We contribute to the literature by focusing on the role of sabotage, which introduces an element of hostile intent absent from standard tournament design and yet pervasive across organizations and labor markets.

The remainder of the paper is structured as follows: Section 2 describes the experimental design, Section 3 reports the results, and Section 4 concludes.

1.2 The Experiment

We conducted six experimental sessions at the Verona Experimental Lab in Economics (VELE), University of Verona. The experiment was computerized using Ztree (Fischbacher, 2007). Recruitment was carried out through direct invitations using ORSEE (Greiner, 2004). We recruited a total of 120 undergraduates, divided into two different treatments, randomized between sessions[2].

[1] Design IDs: HCA40, ICP06, IZU58, MUE79, NCW80, QLM89.

[2] Sessions were scheduled in May 2023. One treatment took place on May 9, 11, and 12 at 14:00, 12:30, and 10:00, respectively. The other treatment was scheduled on May 11 at 10:00 and 15:00, and on May 12 at 12:30.

Each treatment consists of three parts. At the end of the experiment, one of the three parts was randomly selected for payment. Each session lasted 90 minutes, and participants earned on average 17.5 euros. Each candidate was allowed to participate in one session only. Instructions were read aloud, and participants could ask questions if they had any doubts.

In the first part, we elicited one-shot cooperation decisions in both treatments by using a standard linear public goods game (PGG). At the beginning, we read the instructions of the PGG aloud[3]. To ensure that subjects fully understood and learned about the social dilemma situation, participants tested the rules of the game via a simulator for three minutes and then answered a set of control questions (for details, see Appendix A). In the PGG, each subject was assigned to a group of four people and endowed with 20 tokens, which could be either kept for themselves or contributed to a project, the public good. The payoff function is given by:

$$\pi_i = 20 - g_i + 0.5 \sum_{j=1}^{4} g_j \tag{1.1}$$

where π_i is player i's total payoff and g_i his contribution to the public good. The total value of the public good is equal to the sum of the contributions g_j of all group members. Standard assumptions predict that all subjects choose $g_i = 0$. Together with the cooperation decision, we asked subjects to report how much they believe other group members would contribute. Participants did not receive any feedback at this stage.

Part two consists of a fifteen-period tournament. The two treatments differ with respect to the presence of sabotage in the last ten periods of the competition. Hereafter, we will label the treatment without sabotage as "NoSabotage" and the treatment with sabotage as "Sabotage". In both treatments, subjects engaged for five periods in a standard first-price All-Pay Auction (APA) with four players and complete information (Baye et al., 1996; Gneezy and Smorodinsky, 2006; Lugovskyy et al., 2010). Subjects spent two minutes practicing the instructions with a calculator of a one-period APA and answered a set of control questions (see Appendix A for details). Participants were endowed with 1500 tokens to compete with the same group members as in part one. In each period, a prize of 100 tokens was awarded. All bidders' valuations are therefore identical and common knowledge. In a standard APA, each bidder simultaneously submits a sealed bid e_i, which we interpret as the chosen level of effort. The group member with the highest effort wins the prize, and every player must pay the chosen level of effort ($C(e_i) = e_i$). Ties are broken randomly. Therefore, the expected payoff (Contest Success Function) of a risk-neutral agent is:

$$E[\pi_i(\boldsymbol{e})] = \begin{cases} 100 - e_i & \text{if } e_i > \max\{\boldsymbol{e_{-i}}\} \\ \frac{100}{k} - e_i & \text{if } e_i = \max\{\boldsymbol{e}\} \text{ and } k = |j \in N : e_j = \max\{\boldsymbol{e}\}| \\ -e_i & \text{otherwise} \end{cases} \tag{1.2}$$

Theory predicts the existence of a unique symmetric equilibrium and a continuum of asymmetric

[3]Instructions were displayed on screen, participants could not proceed before the appearance of the button "Avanti" (Proceed). We ensured participants engaged with the instructions for a minimum of three minutes. Alongside the text, we presented a visual representation of the instructions, which we report in Appendix A.

equilibria Baye et al. (1996). In each equilibrium, at least two players randomize continuously on the interval [0, 100]. Each other player randomizes continuously on $[e_i, 100]$, where $e_i \geq 0$ is arbitrary, and bids 0 with positive probability if $e_i > 0$. Moreover, the expected payoff of each player is zero, and the expected sum of the bids in any equilibrium is 100. Therefore, from theoretical predictions, we can obtain a measure of efficiency for the competitive environment by observing how much the sum of the bids exceeds the value of the prize in each round.

After five periods, subjects assigned to the "Sabotage" treatment received new instructions. We adopt the formulation used in Harbring and Irlenbusch (2008, 2011) to allow reductions in the probability of winning of the opponents in a first-price all-pay auction. In this framework, the bid of player i is computed as follows:

$$b_i = e_i - \sum_{j \neq i}^{N} s_j \tag{1.3}$$

where e_i is the chosen level of effort and $\sum_{j \neq i}^{N} s_j$ is the sum of the sabotage chosen by the other group members. The marginal cost of effort is equal to 1, while the marginal cost of sabotage is set to be equal to 0.9, reflecting the cost advantage of choosing sabotage over effort[4]. In our case, the expected payoff is:

$$E[\pi_i(\boldsymbol{e}, \boldsymbol{s})] = \begin{cases} 100 - e_i - 0.9s_i & \text{if } b_i > \max\{\boldsymbol{b_{-i}}\} \\ \frac{100}{k} - e_i - 0.9s_i & \text{if } b_i = \max\{\boldsymbol{b}\} \text{ and } k = |j \in N : b_j = \max\{\boldsymbol{b}\}| \\ -e_i - 0.9s_i & \text{otherwise} \end{cases} \tag{1.4}$$

Theoretical predictions imply that participants use only sabotage, where the upper threshold of the interval in which subjects randomize their choice is equal to $\frac{100}{0.9}$. Given the complexity of the game, all participants play the standard APA for the first five periods. In period six, we provide a new set of instructions with the changes in the game's rules. Once again, participants tested the rules of the game through a simulator[5] and answered the related control questions (see Appendix A).

In both treatments, subjects receive feedback. For each round, participants were informed about the chosen level of effort, the winner of the prize, and the residual endowment of each group member. In the "Sabotage" treatment, participants also knew the amount of sabotage received and the value of the final bid.

In part three, participants are re-matched in groups of perfect strangers to play the same public good game as in part one. Feedback on the two PGGs played was given at the end of the experiment. With the current design, we rule out concerns related to reputation, negative reciprocity, and income in the first public good game, allowing us to attribute the treatment effect solely to the spillovers of the competitive environment's features. Furthermore, because participants were familiarized with the instructions through a simulator of the games played,

[4]The treatment manipulation requires the cost function $c(\cdot)$ to satisfy $c(s_i) \leq c(e_i)$. We do not consider the case in which $c(s_i) \geq c(e_i)$, as in that scenario a rational agent would always choose effort.

[5]To view all the information generated by the calculator, subjects engaged with the simulator for four minutes. A representation of the simulator is available in Appendix A (Figure A5).

we limited the impact of potential learning effects. Additionally, since cooperation decisions are one-shot and taken with different group members, we exclude end-game effects. Nevertheless, participants within the same competition group in part two share a common competitive experience, which may induce residual correlation in outcomes. To account for this potential interdependence, we cluster standard errors at the competition-group level in all regression specifications. At the beginning of part three, we read a shorter version of the instructions, in which we informed the subjects of the new group composition. Participants choose how many of the 20 tokens to allocate to the public good. At the end of parts one and three, subjects fill in a questionnaire. In the first questionnaire, we measure trust and fairness levels using standard questions from the European Social Survey (ESS) to explore potential mechanisms behind changes in contributions after exposure to the competitive environment. The final survey includes socio-demographic characteristics, a synthetic measure for the big five personality traits (Chiorri et al., 2015), cognitive reflection test (Frederick, 2005), and participation in social activities. We additionally repeat the trust and fairness questions of part one. Payment is determined based on the results of one of the three parts.

Our main hypothesis is that the amount of achieved cooperation in part three will differ between the two treatments. We expect participants to react adversely to an environment where the probability of success does not depend only on their own choices, especially when others deliberately take actions to reduce the probability of winning of the contestants [6]. Hence, in both contests (with and without sabotage), we expect a reduction of contributions in the public good game played after the tournament. The latter hypotheses were pre-registered using AsPredicted.org[7] before any sessions were conducted.

H_0: Cooperation decreases in both groups after the exposure to a competitive environment (*Competition Hypothesis*).

H_1: Participants assigned to the Sabotage condition will exhibit significantly lower levels of cooperation, as measured by contributions to the second public good, compared to participants in the control condition (*Sabotage Hypothesis*).

In line with studies adopting a similar experimental design, we expect a reduction in prosociality after exposure to a competitive environment. Additionally, we regard our observed effect as a conservative estimate of the actual impact, due to possible anchors established in the first cooperation decision. Furthermore, we assume that the presence of sabotage options leads to a more adversarial and competitive environment. Specifically, we expect that exposure to destructive behaviors (i.e., sabotage) alters both participants' beliefs about the strategies and intentions of others and their own willingness to cooperate. To test our hypotheses, we collect 120 observations (60 per condition). The between-subjects comparison reflects the novel part of the study and is therefore used as reference to compute the minimum detectable effect (MDE).

[6] This statement (point 5 of the pre-registration) is intended to capture two directional hypotheses: (i) a reduction in cooperation following exposure to competitive environments in which outcomes are not solely determined by one's own choices (as in the standard version of the APA), and (ii) a stronger reduction in cooperation when competitors can deliberately sabotage others (reflected in the APA with sabotage). Nevertheless, all statistical tests are conducted using two-tailed specifications, and the minimum detectable effect is calculated accordingly to adopt a more conservative approach.

[7] Pre-registration number AsPredicted #131401: https://aspredicted.org/7f9q-7zz9.pdf.

Assuming a significance level of $\alpha = 0.05$ and a power of $1 - \beta = 0.80$, the MDE corresponds to a Cohen's d of 0.45 under a one-tailed test and 0.51 under a two-tailed test. To adopt a more conservative approach, we rely on the two-tailed specification. Hence, with 60 subjects per treatment, we are powered to detect treatment effects of approximately half a standard deviation in contributions to the public good.

Finally, we report departures from our original pre-analysis plan. The preregistration specifies that the main hypotheses would be tested using a non-parametric Mann–Whitney test to compare contributions across treatments, followed by OLS regressions including the treatment variable, without restrictions on the specification of additional controls. While the empirical strategy broadly follows this plan, a few departures from the preregistered analysis were made. First, for the within-subject comparison of contributions before and after the contest (Competition Hypothesis), we employ a paired-sample sign test instead of a Mann–Whitney test, as the former is more appropriate for within-group comparisons of paired observations. Second, we report additional regression specifications (Columns 3 and 4 of Table 1.1) that were not part of the preregistration, but are included as they offer additional insights into the relationships under study. Third, we present descriptive statistics for behavior in the All-Pay Auction stage, although no analyses of these variables were preregistered. More generally, all statistical tests in the paper that do not concern contributions in the Public Good game were not part of the preregistered analysis plan.

In the next section, we present the results of our experiment. In the first part of the results, we test our two preregistered hypotheses. In the second part, we provide descriptive evidence on the dynamics of the competitive environments, which helps shed light on the mechanisms behind the observed outcomes.

1.3 Results

1.3.1 The Public Good Game

Figure 1.1 displays the average contribution to the public good for each group before and after the exposure to the competitive environment. Participants in the NoSabotage group contribute on average 7.8 tokens before the competition, and 5.9 tokens after the tournament. Subjects in the Sabotage treatment give 8.5 and 6.8 tokens before and after the exposure to the competitive environment, respectively. We observe a decrease in contributions after competing in a contest (p-value$<$0.001, paired-sample sign test). This difference is also observed when considering the two treatments separately (p-values$<$0.001). Beliefs about others' contribution change accordingly: participants predict the decrease in contributions after exposure to the competitive environment (p-value=0.002). The latter result holds for both the analysis of the whole sample and separately for each treatment (p-value=0.049, NoSabotage; p-value=0.012, Sabotage).

Overall, the loss of social surplus can be compared to that found in Buser and Dreber (2016)[8]. In our data, efficiency is 24% lower in the NoSabotage treatment and 20% lower in

[8]The Authors define social surplus "as the difference between the average public goods game payoff and the initial endowment".

the Sabotage treatment. These results are in line with the 19% decrease found by Buser and Dreber (2016) in their experiment, whose experiment reflects the closest design to ours by emphasizing the sequential nature of the competition and cooperation decisions[9]. Furthermore, the study by Li et al. (2023) shows that opponents' trustworthiness mediates the negative effect of competition on cooperative attitudes. In our case, analyzing the data from the questionnaire conducted before and after the game (not pre-registered), trust levels decrease after the contest (p-value=0.002, paired-sample sign test). Furthermore, perceived fairness changes when considering the whole sample (p-value=0.010), but not in the NoSabotage treatment alone (p-value=0.121), suggesting that the treatment manipulation was effective in generating higher levels of perceived unfairness. Altogether, the exclusion of the mechanisms previously mentioned and changes in beliefs and trust levels support the existing evidence on the negative effect of competition on cooperation.

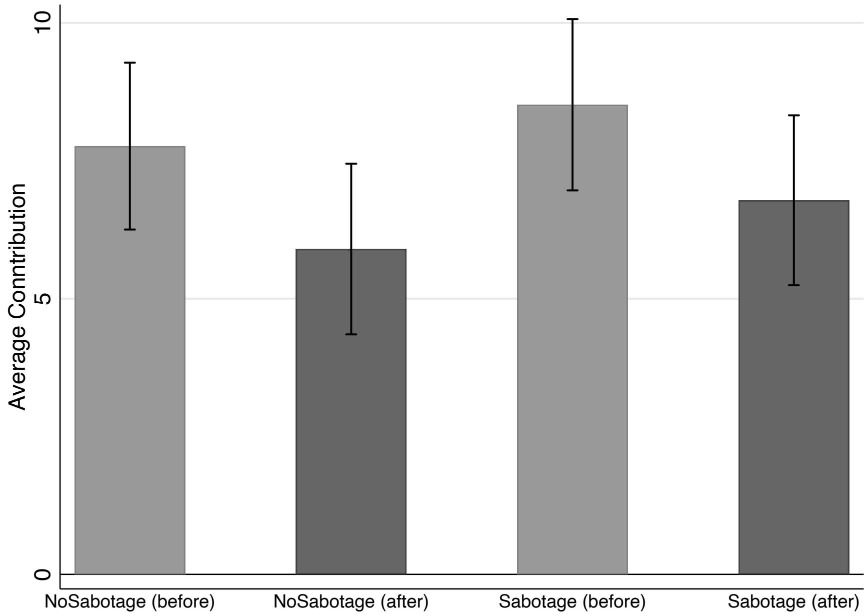


Figure 1.1: Contribution before and after the exposure to the competitive environment. Error bars display the 95% confidence interval.

Next, we investigate whether competition with sabotage (with respect to standard competition) has a negative impact on cooperation. Comparing the level of achieved cooperation in the second PGG across treatments, we find that sabotage in competition does not significantly affect cooperation (p-value = 0.303, Mann-Whitney test). In our experimental setting, the reduction in contributions is slightly larger in the NoSabotage treatment (9.5% versus 8.5% of the initial endowment). We can therefore conclude that, within our experimental setting, the treatment variation does not affect cooperative attitudes.

We further explore the role of sabotage on cooperation in the second PGG using ordinary least squares (OLS) regressions with clustered standard errors at the competition group level[10].

[9] In their experiment, participants are paid based on the number of sliders they solve under both a piece-rate and a competitive payment scheme, and they play the PGG after being reassigned to new groups.

[10] Although subjects are re-matched into groups of perfect strangers in part three, members of the same competition group may exhibit residual correlated outcomes due to their common prior exposure to the tournament.

In Table 1.1 we regress contributions in the second PGG over the treatment variable "Sabotage", beliefs on others' contributions, beliefs and contributions in the first PGG (column (1)), and a set of control variables (column (2)). We find no evidence of differing contributions among subjects assigned to the two treatments (column (1), p-value = 0.485). This result is robust to the inclusion of observables (column (2), p-value = 0.784). Consistent with the existing literature, cooperation is primarily driven by beliefs regarding others' contributions. Beliefs in the first PGG do not predict contributions in the second PGG, suggesting that participants update their beliefs after exposure to the competitive environment. Moreover, higher response time is associated with lower cooperation, and being a member of different associations ("Social Activities") positively correlates with cooperative attitudes.

Finally, in columns (3) and (4), we study how the average amount of received sabotage affects cooperation using the whole sample and within the Sabotage group only. The latter analysis was not preregistered; we therefore include models (3) and (4) as exploratory analyses. The amount of sabotage received is one of the feedback that participants receive in each round and might prompt participants negatively, leading to stronger reactions when deciding their own contribution. We find that subjects who received more sabotage contribute less in the PGG played after the tournament. Column (3) reports the results using the whole sample, where the received sabotage is equal to zero in the control group by construction (column (3), p-value = 0.010). As a robustness test, we repeat the analysis for the subsample of treated subjects (Column (4)), confirming that the amount of received sabotage negatively correlates with contributions in the public good game (column (4), p-value = 0.006). A one-standard-deviation increase in the received sabotage leads to a reduction of 0.15 standard deviations in contribution to the public good[11]. The latter finding suggests that treatment intensity plays a role in our context, as the greater use of sabotage may contribute to a stronger decline in subsequent prosocial behavior.

1.3.2 The All-Pay Auction

In this section, we provide a brief description of the choices made by participants during the competition. We further discuss the role of sabotage in the All-Pay Auction (APA) framework and its efficiency implications[12].

Figure 1.2 (Panel A) displays the average levels of chosen effort and sabotage across the two treatments. At the beginning of the competition, the choices reflect a pronounced tendency toward excessive overbidding, which rapidly decreases after period 2. Following this initial phase of learning, participants in the Sabotage treatment simultaneously determine their levels of effort and sabotage. The gradual escalation in the use of sabotage over time suggests a progressive substitution of effort with sabotage by the participants. However, the mechanisms that underlie the substitutability between effort and sabotage are unclear. Participants may engage in reciprocal actions in response to the sabotage received, they might recognize the economic advantages of selecting sabotage over effort, or they could be systematically experimenting with

[11] This estimate should be interpreted with caution, as our original a priori power analysis was designed to detect a larger effect size of 0.5 standard deviations with the collected sample.

[12] Figures and analyses in this section were not preregistered.

	(1) Contribution	(2) Contribution	(3) Contribution	(4) Contribution
Sabotage	0.05	0.02	0.11*	
	(0.07)	(0.06)	(0.06)	
Contribution (Pre)	0.68***	0.66***	0.67***	0.81***
	(0.08)	(0.08)	(0.08)	(0.10)
Belief (Pre)	-0.13	-0.16	-0.16	-0.12
	(0.12)	(0.10)	(0.10)	(0.14)
Belief (Post)	0.33***	0.32***	0.32***	0.16*
	(0.10)	(0.08)	(0.08)	(0.09)
Trust (Post)		0.08	0.08	0.04
		(0.07)	(0.06)	(0.06)
Fairness (Post)		0.09	0.10	0.09
		(0.06)	(0.06)	(0.08)
Risk Preference		-0.05	-0.06	-0.10
		(0.04)	(0.04)	(0.06)
Response Time		-0.16***	-0.15***	-0.15***
		(0.06)	(0.06)	(0.04)
Religiosity		0.07	0.08*	0.17**
		(0.05)	(0.05)	(0.06)
Female		-0.07	-0.07	-0.12*
		(0.05)	(0.05)	(0.06)
Social Activities		0.12**	0.11**	0.06
		(0.05)	(0.05)	(0.06)
Year of study		-0.11*	-0.12*	-0.15**
		(0.06)	(0.06)	(0.06)
Received Sabotage			-0.13***	-0.15***
			(0.05)	(0.05)
Constant	-0.00	-0.00	-0.00	0.12**
	(0.06)	(0.06)	(0.06)	(0.05)
Observations	120	120	120	60
R^2	0.649	0.720	0.726	0.831

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

Table 1.1: OLS estimates

different approaches while adjusting their choices at each stage based on the feedback received.

The chosen level of effort and sabotage determines the value of the final bid in the Sabotage treatment. Figure 1.2 (Panel B) shows the average value of the final bid in the two groups. In the NoSabotage treatment, final bids are exactly equal to the level of chosen effort (Panel A), while effort levels are reduced by the sabotage chosen by the other group members in the Sabotage treatment (Harbring and Irlenbusch, 2008, 2011). Before the new set of rules was given to subjects assigned to the Sabotage group, participants of different treatments behaved similarly. Following period 5, bids in the Sabotage group are reduced substantially, indicating that the treatment manipulation successfully generates positive levels of sabotage.

Efficiency in APAs usually refers to the notion of over-dissipation, which occurs when the sum of all group members' bids exceeds the value of the prize (Baye et al., 1996). Figure 1.3 (Panel A) shows the average total cost in each treatment at each round of the competition. Despite subjects in Sabotage treatment having the possibility to save tokens by substituting sabotage with effort, they generally spend more resources. We obtain a measure of over-dissipation in this context by summing all the costs incurred by each group at every round and subtracting the value of the prize. Results from the Mann–Whitney suggest that over-dissipation is larger

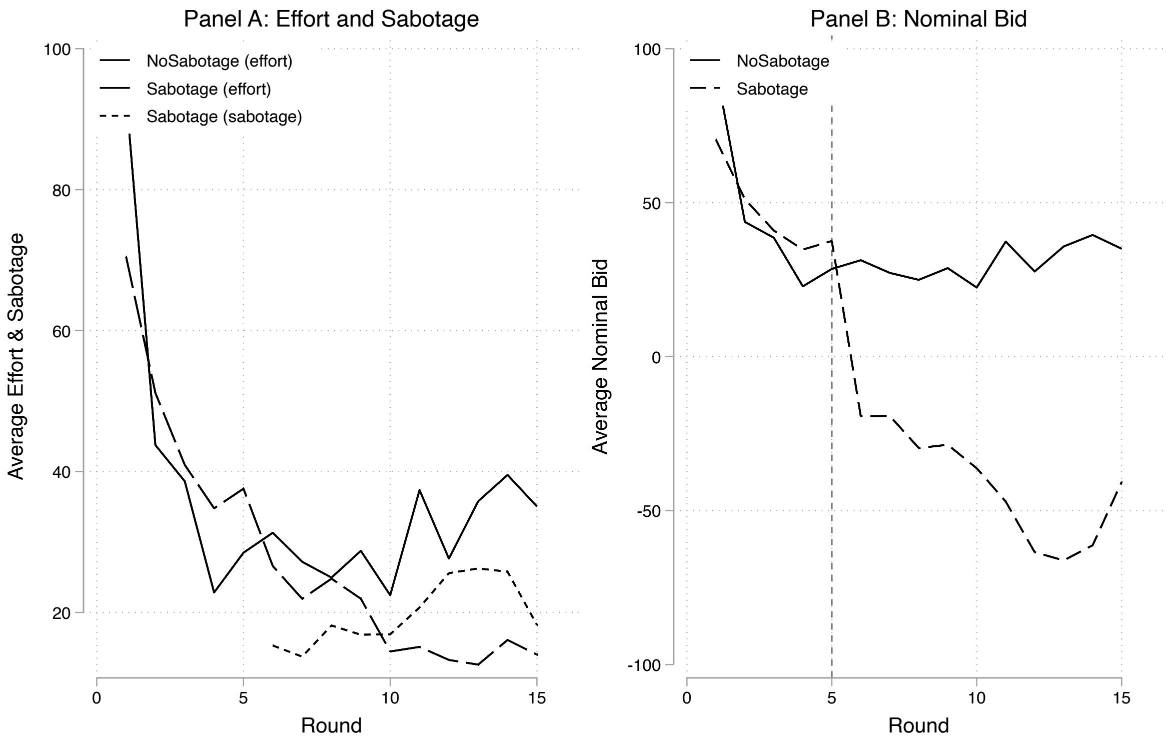


Figure 1.2: Panel A shows chosen effort and sabotage. The solid line displays the average effort level (bid) chosen by participants in the NoSabotage treatment (i.e., the standard APA). The dashed lines refer to the average chosen effort (long-dash) and chosen sabotage (short-dash) in the Sabotage treatment. Panel B displays the average nominal bids. The solid line shows the average bids in the NoSabotage treatment, while the dashed line shows the final bids in the Sabotage treatment, computed using the formulation of Harbring and Irlenbusch (2011).

in the Sabotage treatment (p-value=0.024). Figure 1.3 (Panel B) shows that costs are larger in the Sabotage group from round 4 to round 10. Over-dissipation may therefore be due to the adjustments that participants make while understanding how to fully exploit their options. An alternative explanation is that participants in the NoSabotage treatment raise their bid during the last periods, in response to a low-bidding behavior in the middle part of the competition.

To further evaluate efficiency concerns of the competition, we explore the difference in the final endowment. Figure 1.3 (Panel B) shows the trend of the average endowment for each treatment. We expect that if participants systematically overbid in the Sabotage treatment, then the two groups' endowments result different in the last period of the competition. Using the parametric t-test, we found no significant differences in the final endowment between the two groups (p-value=0.3043). While the direction of the effect of sabotage on efficiency is clear, the number of observations may not be sufficiently high to conclude that efficiency represents a concern in this setting.

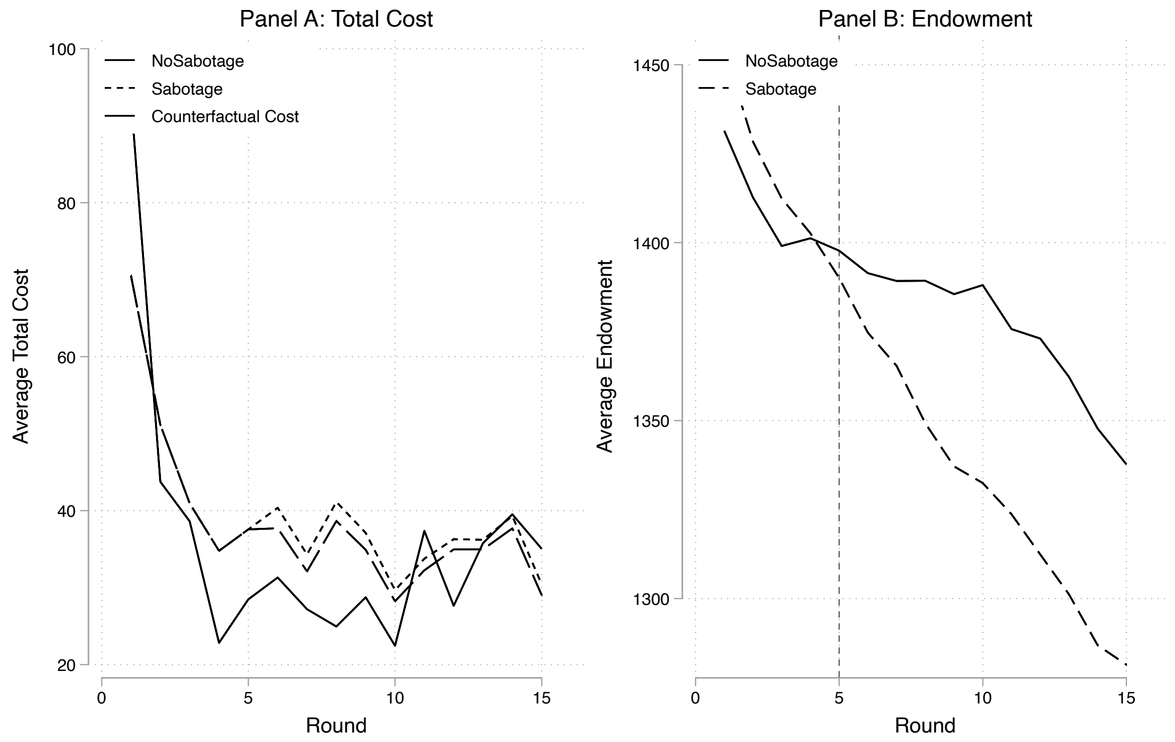


Figure 1.3: Panel A shows the average total cost incurred by participants in the NoSabotage (solid line) and Sabotage (short-dash line) treatments. It also includes a counterfactual cost (long-dash line) representing what would have been the cost of bidding in the Sabotage treatment if participants had used only sabotage. Panel B shows the average endowment per round in both treatments. The solid line refers to the NoSabotage treatment, while the dashed line shows the corresponding values in the Sabotage treatment.

1.4 Conclusion

In this study, we analyze how exposure to competitive environments affects cooperation decisions in two respects. First, we observe changes in cooperation before and after participants engage in a competition. In line with previous studies with a similar experimental design (Brandts and Riedl, 2020; Buser and Dreber, 2016; Li et al., 2023), we provide evidence of the existence of negative spillover effects of competition on prosocial behavior. Nevertheless, this result stands in contrast with the findings of Chapter 2, which replicates a similar research question in an online setting using a real-effort task as the competitive stage. In Chapter 2, no significant detrimental effect of tournament incentives on public good contributions is detected, a result consistent with the heterogeneity of different experimental designs documented by Huber et al. (2023).

Second, we investigate the causal effect of sabotage in competitions on subsequent prosocial behavior. Our design allows us to interpret the result as a mere psychological spillover stemming from the presence of destructive activities within the competitive environment. We find no significant effect of sabotage on contributions in a randomized setting. However, when focusing solely on the Sabotage condition, contributions decrease with the amount of sabotage received during the competition, indicating that higher treatment intensity can be more effective in exploring the consequences of sabotage actions.

The absence of treatment differences between the standard and unfair competitions suggests that simply introducing the possibility of destructive actions within the competition does not lead to disruptive changes in the existing negative spillovers on cooperation (MDE = 0.5 s.d.). One interpretation is that participants do not react to reductions in their probability of winning and spillovers are not affected beyond the standard competitive setting. Another interpretation is that when sabotage is framed as a legitimate component of the competitive environment, participants may normalize destructive behavior, reducing its psychological impact on subsequent prosocial choices. Lastly, destructive actions may have stronger behavioral spillovers when they cause real inefficiencies rather than only affecting the probability of success.

Future research on the potential drawbacks of tournament incentives can provide important insights. The literature still lacks a robust framework for understanding how competition affects cooperation in laboratory and online experiments (Huber et al., 2023). Different contest structures (stakes, repeated interactions, real-effort tasks) often lead to divergent conclusions about the existence and direction of spillover effects. Field experiments, while less common (Kosse et al., 2025), could help clarify whether the effects of competition generalize to natural settings where stakes, norms, and social relationships are not artificially constructed. Another important aspect concerns the role of sabotage: although destructive behavior is widespread in many competitive environments, its efficiency costs and its implications for selection into competitive settings remain insufficiently understood.

The presence of negative spillovers from competition to cooperation has relevant implications for organizations and educational institutions. Because tournament incentives can boost performance but lower cooperation, decision-makers face a trade-off between rewarding relative performance and preserving cooperative norms. Organizations may find it beneficial to use competitive schemes when tasks depend on individual performance, but rely more heavily on team incentives or group-based rewards when cooperation is an essential aspect of the core business (Markussen et al., 2014). Similarly in education, if schools aim to increase their reputation by selecting high-skill students, then competition can potentially result in better performance, but at the risk of sabotage activities. If otherwise, the goal is to reduce inequalities among low and high-skill students, then fostering cooperation can be optimal.

Ultimately, it is crucial to consider the welfare implications of the relationship presented in this paper. If competition leads to a decrease in trust and fairness levels, negative spillover effects can affect outcomes beyond educational institutions and workplaces, subsequently diminishing the overall quality of social interactions and welfare.

References

Angelovski, A., Neugebauer, T., and Servátka, M. (2019). Rank-order competition in the voluntary provision of impure public goods. *Economic Inquiry*, 57(4):2163–2183.

Augenblick, N. and Cunha, J. M. (2015). Competition and cooperation in a public goods game: A field experiment. *Economic inquiry*, 53(1):574–588.

Baye, M. R., Kovenock, D., and De Vries, C. G. (1996). The all-pay auction with complete information. *Economic Theory*, 8:291–305.

Bergantino, A. S., Gil-Gallen, S., and Morone, A. (2021). Do risk and competition trigger conditional cooperation? evidence from public good experiments. *Annals of Public and Cooperative Economics*.

Brandts, J. and Riedl, A. (2020). Market interaction and efficient cooperation. *European Economic Review*, 121:103318.

Buser, T., Cappelen, A. W., and Tungodden, B. (2021). Fairness and willingness to compete. *NHH Dept. of Economics Discussion Paper*, N.A.(08).

Buser, T. and Dreber, A. (2016). The flipside of comparative payment schemes. *Management Science*, 62(9):2626–2638.

Buser, T. and Sangi, S. (2025). Willingness to compete in dirty competitions.

Canegallo, C., Ortona, G., Ottone, S., Ponzano, F., and Scacciati, F. (2008). Competition versus cooperation: Some experimental evidence. *The Journal of socio-economics*, 37(1):18–30.

Cason, T. N. and Gangadharan, L. (2013). Cooperation spillovers and price competition in experimental markets. *Economic Inquiry*, 51(3):1715–1730.

Charness, G., Masclet, D., and Villeval, M. C. (2014). The dark side of competition for status. *Management Science*, 60(1):38–55.

Chen, K.-P. (2003). Sabotage in promotion tournaments. *Journal of Law, Economics, and Organization*, 19(1):119–140.

Chiorri, C., Bracco, F., Piccinno, T., Modafferi, C., and Battini, V. (2015). Psychometric properties of a revised version of the ten item personality inventory. *European Journal of Psychological Assessment*.

Colasante, A., García-Gallego, A., Georgantzis, N., Morone, A., and Temerario, T. (2019). Intragroup competition in public good games: The role of relative performance incentives and risk attitudes. *Journal of Public Economic Theory*, 21(5):847–865.

Colasante, A., García-Gallego, A., Morone, A., and Temerario, T. (2017). The utopia of cooperation: does intra-group competition drive out free riding. *Working Papers*, 8.

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

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, 19(4):25–42.

Gneezy, U. and Smorodinsky, R. (2006). All-pay auctions—an experimental study. *Journal of Economic Behavior & Organization*, 61(2):255–275.

Godoy, S., Morales, A. J., and Rodero, J. (2013). Competition lessens competition: An experimental investigation of simultaneous participation in a public good game and a lottery contest game with shared endowment. *Economics Letters*, 120(3):419–423.

Greiner, B. (2004). An online recruitment system for economic experiments.

Gürtler, O., Münster, J., and Nieken, P. (2013). Information policy in tournaments with sabotage. *The Scandinavian Journal of Economics*, 115(3):932–966.

Harbring, C. and Irlenbusch, B. (2008). How many winners are good to have?: On tournaments with sabotage. *Journal of Economic Behavior & Organization*, 65(3-4):682–702.

Harbring, C. and Irlenbusch, B. (2011). Sabotage in tournaments: Evidence from a laboratory experiment. *Management Science*, 57(4):611–627.

Huber, C., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Weitzel, U., Abellán, M., Adayeva, X., Ay, F. C., Barron, K., et al. (2023). Competition and moral behavior: A meta-analysis of forty-five crowd-sourced experimental designs. *Proceedings of the National Academy of Sciences*, 120(23):e2215572120.

Konrad, K. A. (2000). Sabotage in rent-seeking contests. *Journal of Law, Economics, and Organization*, 16(1):155–165.

Kosfeld, M. and Von Siemens, F. A. (2011). Competition, cooperation, and corporate culture. *The RAND Journal of Economics*, 42(1):23–43.

Kosse, F., Rajan, R., and Tincani, M. (2025). The persistent effect of competition on prosociality. *Journal of the European Economic Association.*

Lazear, E. P. (1989). Pay equality and industrial politics. *Journal of political economy*, 97(3):561–580.

Li, Y., Liu, Z., Wang, Y., Derrington, E., Moisan, F., and Dreher, J.-C. (2023). Spillover effects of competition outcome on future risky cooperation. *Scientific Reports*, 13(1):5535.

Lien, J. W., Zheng, J., and Zhuo, Y. (2021). The cooperative consequences of contests. *Available at SSRN 3850570.*

Lugovskyy, V., Puzzello, D., and Tucker, S. (2010). An experimental investigation of overdissipation in the all pay auction. *European Economic Review*, 54(8):974–997.

Markussen, T., Reuben, E., and Tyran, J.-R. (2014). Competition, cooperation and collective choice. *The Economic Journal*, 124(574):F163–F195.

Moyal, A. and Ritov, I. (2020). The effect of contest participation and contest outcome on subsequent prosocial behavior. *Plos one*, 15(11):e0240712.

Riley, J. G. and Samuelson, W. F. (1981). Optimal auctions. *The American Economic Review*, 71(3):381–392.

Savikhin, A. C. and Sheremeta, R. M. (2013). Simultaneous decision-making in competitive and cooperative environments. *Economic Inquiry*, 51(2):1311–1323.

Thöni, C., Tyran, J.-R., and Wengström, E. (2012). Microfoundations of social capital. *Journal of Public Economics*, 96(7-8):635–643.

Appendix A

In this section, we report the instructions, the control questions and the simulators used in each part of the original experiment. Instructions were displayed on each participant's monitor and read aloud. We made use of figures to ease the understanding of the rules. The latter are attached at the bottom of this section.

English version

A.1 Instructions

General Rules Welcome! Thank you for agreeing to participate in the experiment. For your participation, you will receive 3 euros, to which an additional amount may be added based on the decisions made during the experiment. Throughout the experiment, it is prohibited to talk to any other participant. If you have any questions, please raise your hand and wait for a staff member to respond to you privately. We ask you to turn off your cellphone. If you violate these rules, we will have to exclude you from the experiment, and you will forfeit any payment. Today's experiment consists of three parts. In each part, you will earn tokens, which will be converted into euros at the end of the experiment at a rate that will be communicated to you before the start of each part. At the end of the experiment, one of the three parts will be randomly selected, and your compensation for today's experiment will depend on the tokens earned in that part. Since you do not know which part will be selected for payment, it is important to treat all parts as potentially relevant. Click on CONTINUE to proceed to the next screen.

Part I - Instructions Public Good game

The conversion rate to euros for PART I is: 1 token = 50 cents.

You are together with three other people in this room to form a group of four members. Each group member has 20 tokens in their private account. You can keep these 20 tokens in your private account or contribute them fully or partially to a project.

TOKENS FROM THE PROJECT. The total number of tokens contributed to the project will be doubled and divided equally among the four group members.

Example. If the four group members contribute a total of 60 tokens to the project, then you and the other members of your group will each earn (60 x 2) / 4 = 120 / 4 = 30 tokens from the project.

YOUR TOTAL EARNINGS = TOKENS IN YOUR PRIVATE ACCOUNT + TOKENS FROM THE PROJECT

Image Example. Red, Green, Black, and Blue are four members of a group with 20 tokens in their private accounts (Figure 1). Each of them chooses how much to contribute to the project: Red contributes 0 tokens, Green contributes 15 tokens, Black contributes 5 tokens, and Blue contributes 20 tokens (Figure 2). The total number of tokens contributed to the project is doubled (Figure 3). The tokens in the project are divided into 4 equal parts (Figure 4) and distributed to the group members (Figure 5). The tokens earned from the project are added to those held in the private account. Red has 40 tokens, Green has 25 tokens, Black has 35 tokens, and Blue has 20 tokens (Figure 6). In the next screen, you will have the opportunity to test the functioning of the rules through a simulator. Then, we will ask you some control questions to help you make your decisions. Click on CONTINUE to proceed to the next screen (the button will appear in a few moments).

Part II – Instructions standard All Pay Auction

The conversion rate to euros for PART II is: 100 tokens = 1 euro.

In this part, you will be in a group with the same three people from PART I. You will interact with them for 15 rounds. At the beginning of the first round, each group member will have a credit of 1500 tokens. In each round, the group members will compete to win a prize of 100 tokens. Each member will choose an amount to invest in order to win the prize. The group member who chooses the highest investment will win the prize of 100 tokens. The other three group members will receive nothing. If two or more group members choose the same highest investment, one of them will be randomly selected as the winner of the prize.

Your earnings. Your total earnings will be the remaining credit (in tokens) at the end of the 15th round. At the end of each round, the eventual prize won will be added to your initial credit, and the tokens you invested will be subtracted. Your remaining credit at the end of each round will be calculated as follows:

Your remaining credit at the end of the round = remaining credit from the previous round + eventual prize - investment

NOTE: Your credit cannot go below zero. If your credit reaches zero, you will not be able to participate in the subsequent rounds, and your final earnings will be 0. In the next screen, you will have the opportunity to test the functioning of the rules through a simulator. Then, we will ask you some control questions to help you make your decisions. Click on “CONTINUE” to proceed to the next screen (the button will appear in a few moments).

Part II – Instructions All Pay Auctions with Sabotage

For the next 10 rounds, in addition to the investment decision, you will also be able to reduce the investments of all other group members. Therefore, each member will choose both the amount invested and an amount to reduce the investments of others. The final investment of each member will be their chosen investment minus the sum of the reductions decided by the other group members. For example, if you choose to invest 50 tokens and the others decide to reduce a total of 35 tokens, your final investment will be 15 tokens. The group member with the highest final investment will win the prize of 100 tokens.

Your earnings. Your total earnings will be the remaining credit (in tokens) at the end of the 15th round. At the end of each round, the eventual prize won will be added to your initial credit, and the costs for your choices will be subtracted. This includes both the tokens you invested and the cost for reducing the investments of others. Your remaining credit at the end of each round will be calculated as follows: Your remaining credit at the end of the round = remaining credit from the previous round + eventual prize - investment - reduction cost. For each reduction point you choose to assign to other group members, you will incur a cost of 0.9 tokens. For example, the cost of choosing a reduction of 50 tokens is: 50 x 0.9 = 45 tokens. A button on the decision screen will help you visualize the cost of the two choices. In the next screen, you will have the opportunity to test the functioning of the rules through a simulator. Then, we will ask you some control questions to help you make your decisions. Click on "CONTINUE" to proceed to the next screen (the button will appear in a few moments).

Part III - Public Good game

The conversion rate to euros for PART III is: 1 token = 50 cents.

Like before, you are in a group of 4 people. However, the composition of the group is entirely new. None of the participants who were in your group in the first and second parts of the experiment will be in your group for this part. The instructions are the same as in the first part. Each group member has 20 tokens in their private account. You can keep these 20 tokens in your private account or contribute them entirely or partially to a project. Your earnings will be determined in the same way as before.

THE PROJECT TOKENS. The sum of the tokens contributed to the project will be doubled and divided equally among the four group members. Example: If the four group members contribute a total of 60 tokens to the project, then you and the other members of your group will each earn (60 x 2) / 4 = 120 / 4 = 30 tokens from the project.

YOUR TOTAL EARNINGS = TOKENS IN YOUR PRIVATE ACCOUNT + TOKENS FROM THE PROJECT Click on "CONTINUE" to proceed to the next screen (the button will appear in a few moments).

A.2 Control Questions

Control question Public Good game

Please solve the following independent cases. They will help you understand how your earnings and the earnings of others are calculated in tokens.

Case 1. None of the four group members (including yourself) contribute to the project. So each member keeps all 20 tokens in their private account. Questions:
How many tokens will you have in total?
How many tokens will each of the other group members have in total?

Case 2. All four group members (including yourself) contribute 20 tokens to the project. So each member keeps 0 tokens in their private account. The sum of tokens contributed to the project is 20 + 20 + 20 + 20 = 80. Questions:
How many tokens will you have in total?
How many tokens will each of the other group members have in total?

Case 3. You contribute 20 tokens to the project. None of the other three group members contribute to the project. So you keep 0 tokens in your private account, and each of the other members keeps all 20 tokens in their private account. The sum of tokens contributed to the

project is 20. Questions:
How many tokens will you have in total?
How many tokens will each of the other group members have in total?

Case 4. You contribute 0 tokens to the project. The other three group members contribute 20 tokens each to the project. So you keep 20 tokens in your private account, and each of the other members keeps 0 tokens in their private account. The sum of tokens contributed to the project is 20 + 20 + 20 = 60. Questions:
How many tokens will you have in total?
How many tokens will each of the other group members have in total?

Control questions standard All-Pay-Auction

1. Suppose you invest 20 tokens and do not win. How many tokens will you earn in this round? Answers:
a) You lose 20 tokens,
b) You earn 80 tokens,
c) You lose 120 tokens,
d) You lose 100 tokens.

2. Suppose you invest 0 tokens and do not win. How many tokens will you earn in this round?
a) You lose 100 tokens,
b) You earn 100 tokens,
c) You neither gain nor lose,
d) Offering 0 is not allowed.

3. Suppose in the first round you invest 40 tokens and you win the prize. Answers:
a) Your credit at the end of the round is 1590,
b) Your credit at the end of the round is 1560,
c) Your credit at the end of the round is 1490,
d) Your credit at the end of the round is 1410.

4. Suppose in the first round you invest 110 tokens and you win the prize. Answers:
a) Your credit at the end of the round is 1600,
b) Your credit at the end of the round is 1490,
c) Your credit at the end of the round is 1400,
d) None of the above.

Control questions All-Pay-Auction with Sabotage

1. If the other three members of the group choose a reduction of 20 each. Answers:
a) Your remaining credit decreases by 60,
b) Your final investment is reduced by 20,
c) Your final investment is reduced by 60.

2. If you choose a reduction of 35. Answers:

a) The final investment of only one member is reduced by 35,
b) The final investment of all members is reduced by 35,
c) The remaining credit of all members decreases by 35.

3. The total cost (investment and cost for reduction) that will be subtracted from your credit. Answers:
a) It depends on the choices of others,
b) It depends only on your choices,
c) All answers are correct.

A.3 Images and simulators

In this section we provide the images related to the Public Good game instructions and the simulators used in the experiment. All simulators were played after the instructions.

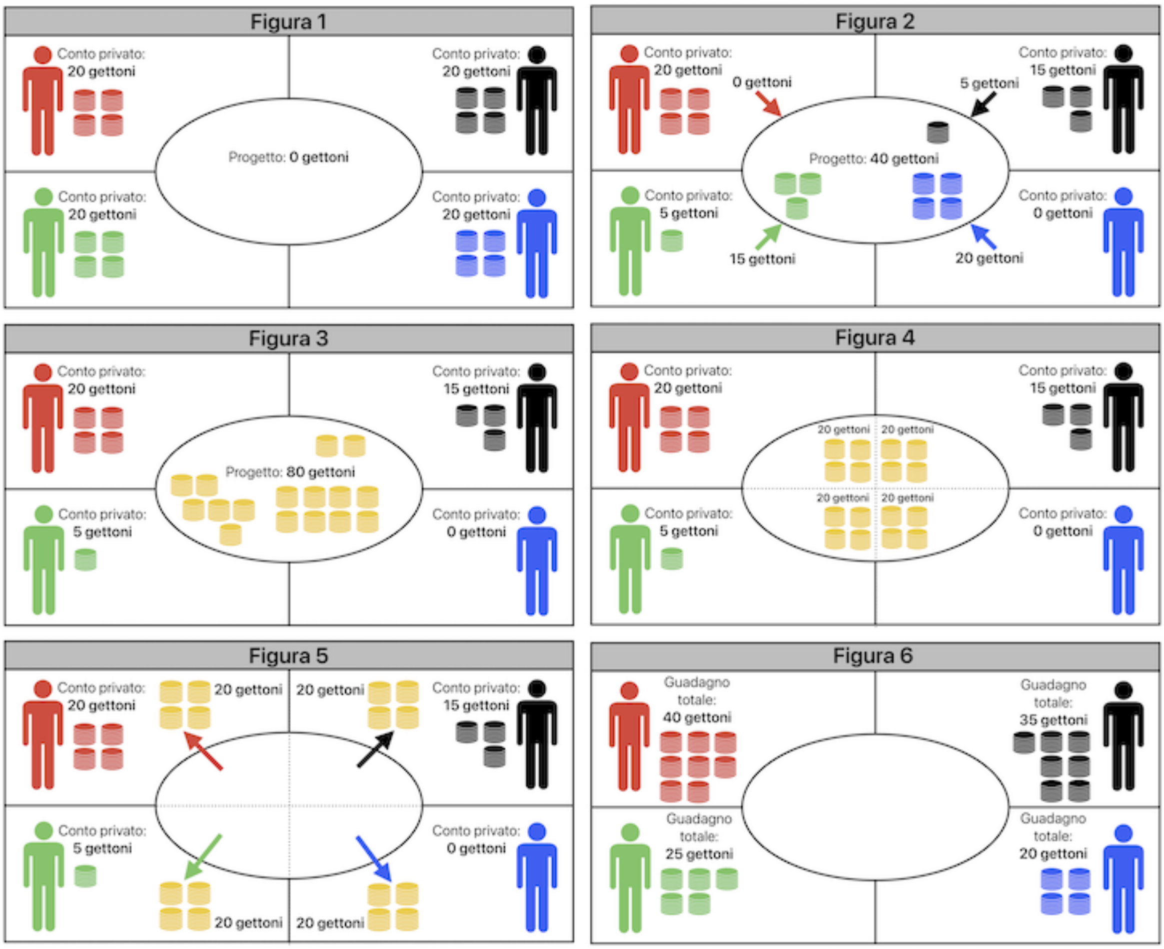


Figure A1: Representation of the Public Good game instructions. Panel (Figura 1) shows the four members of a group, each endowed with 20 tokens. Panel 2 shows the decision situation: each member decides how much to contribute in the project. Panel 3 shows the doubling of the invested tokens. Panel 4 shows the equal subdivision of the tokens in the project. Panel 5 shows the received earnings from the project. Panel 6 shows the total earnings from the game. This representation is inspired by (Thöni et al., 2012).

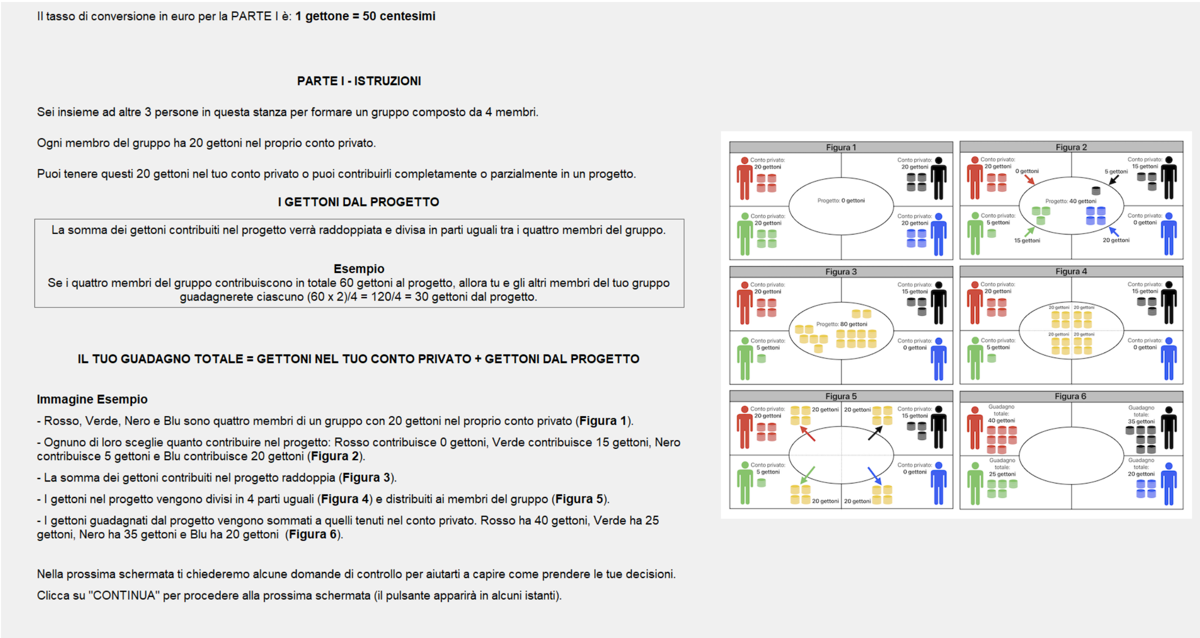


Figure A2: The Public Good game instructions on display.

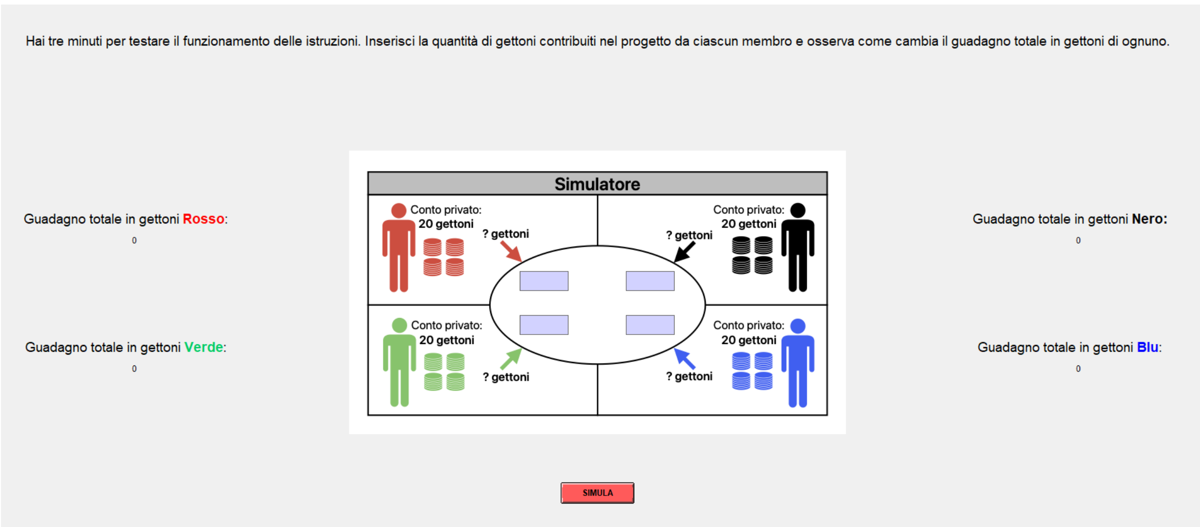


Figure A3: Simulator Public Good game. Participants practice the rules of the game by choosing how much each member contribute in a hypothetical scenario. Once the button "Simula" ("Simulate") is pressed, a player is able to visualise the earnings of each member ("Guadagno totale in gettoni").

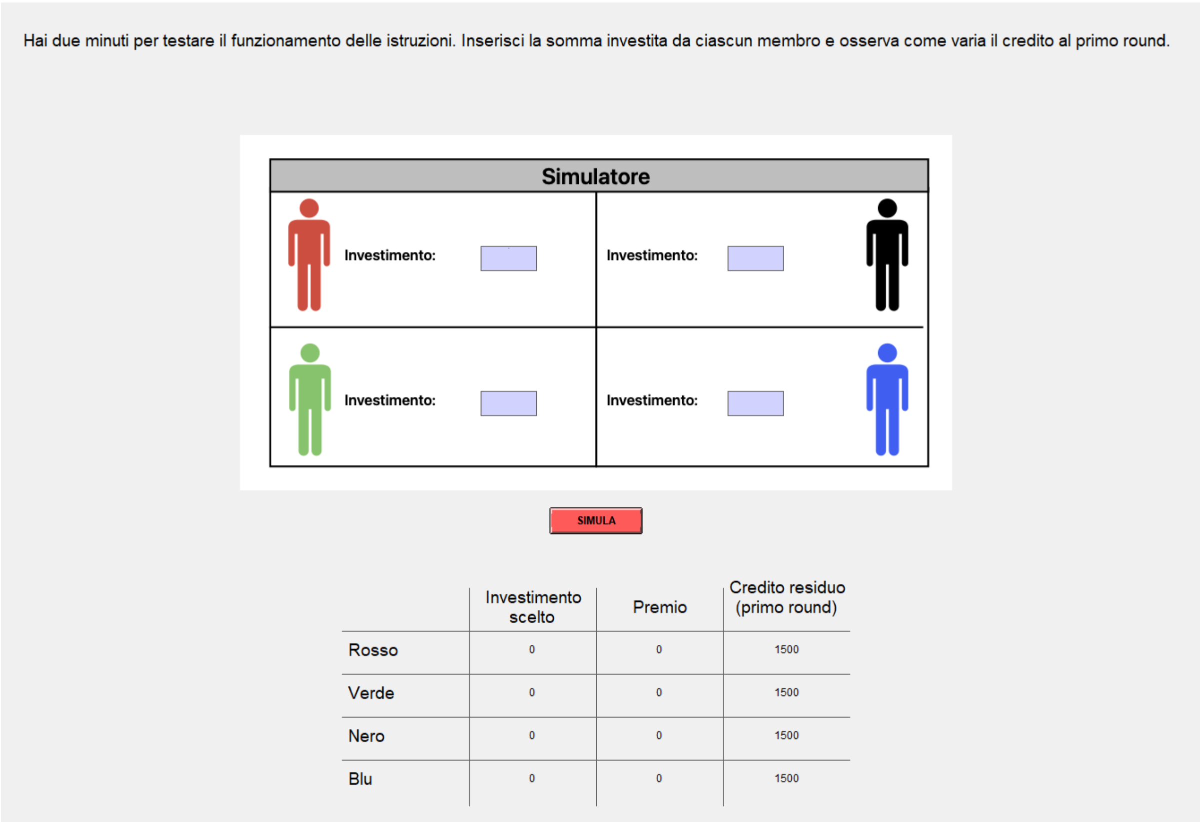


Figure A4: Simulator All-Pay-Auction "NoSabotage". Participants practice the rules of the game by choosing a level of effort for each participant. The table below updates once the button "Simula" ("Simulate") is pressed. The table reports the chosen level of effort, the value of the prize won (either 0 or 100) and the residual credit at the end of the first round. Due to the variations in how the word "effort" is used in the Italian language compared to English, the instructions were not expressed as "chosen effort" but rather as "Investment" to ease the understanding of the game.

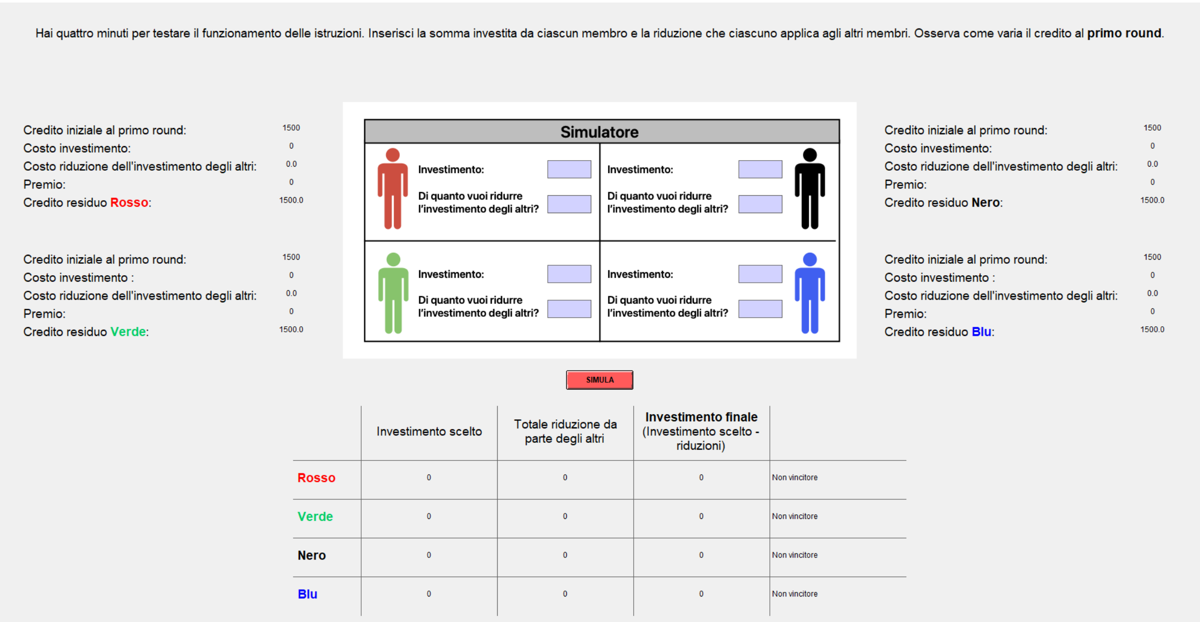


Figure A5: Simulator All-Pay-Auction “Sabotage”. Participants practice the rules of the game by choosing a level of effort and sabotage for each participant. The table below updates once the button “Simula” (“Simulate”) is pressed. The table reports the chosen level of effort, the amount of received sabotage and the value of the final bid, together with a label indicating whether a member is not a winner (“Non vincitore”) or a winner (“Vincitore”). On a side of each member cell, we reported the costs of effort and sabotage and the value of the prize won (either 0 or 100), together with the residual credit. Due to the variations in how the word “effort” is used in the Italian language compared to English, the instructions were not expressed as “chosen effort” but rather as “Investment” to ease the understanding of the game. Similarly, “sabotage” is rather express as a question of the kind “How much you want to reduce the level of investment of the other group members?”

Appendix B

B.1 Pre-Registration

Title: Competition and Cooperation (#131401)
Pre-registered on: 05/08/2023 05:50 PM (PT)
This pre-registration is currently anonymous to enable blind peer-review. It has one author.

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
The goal is to quantify changes in cooperation levels due to exposure to different competitive environments. We test whether a competitive environment allowing sabotage activities decrease subsequent cooperation in a Public Good game compared to an environment where sabotage is not possible.

3) Describe the key dependent variable(s) specifying how they will be measured.
In the first and third part of the experiment, we elicit subjects' willingness to cooperate using a standard one-shot Public Good game. The key dependent variable is the contribution to the Public Good. In the second part, we let subjects compete in a 15-rounds All-Pay-Auction (APA). The variables of interest are the chosen effort and sabotage levels, which are substitutes in the contest success function. The variables' values are directly chosen by participants.

4) How many and which conditions will participants be assigned to?
Each participant will only be assigned to one out of two experimental treatments. The two experimental conditions differ with regard to the rules in the APA contests. In the first condition, participants compete according to standard APA rules. In the second condition, participants are allowed to reduce the probability of success of others (APA with sabotage).

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
Our main hypothesis is that the amount of achieved cooperation in the third part of the experiment will differ systematically between the two treatments. We expect participants to adversely react to environment where the probability of success does not depend only by their own choices, especially when others deliberately take actions to reduce the probability of winning of the contestants. Hence, in both contests (with and without sabotage) we expect a reduction of contributions in the Public Good game played after the tournament. Our main analysis involves the use of a non-parametric test (Mann-Whitney test) and ordinary least square regression to study differences in contributions in the two Public Good games.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
All data from the experiment will be analysed and no observation will be excluded.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We plan to recruit a total of n = 120 participants, 60 for each of the two experimental condition.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

Nothing else to pre-register.

Chapter 2

Restoring a Cooperative Mindset in Competitive Environments: the Role of Donation

A. Barbazeni

Abstract

We study the impact of donation programs on sustaining cooperation after exposure to a competitive environment. We replicate experimental conditions to study the impact of tournament incentives on willingness to cooperate found in Buser and Dreber (2016) and introduce variations in which competition winners can publicly donate to a charity. Our analysis informs on the effectiveness of donation programs in mitigating the potential negative spillover effects of competition on cooperation. We find that competition does not affect cooperation, resulting in the replication being unsuccessful. Furthermore, donation decisions made after a competition reflect winners' prosocial behavior, but they do not change the cooperation of competition losers.

Keywords: Competition; Cooperation; Donation; Signaling.
JEL Codes: C92, D64, M52.

2.1 Introduction

In modern workplaces, the adoption of horizontal hierarchies emphasizes greater delegation of responsibilities and simplified decision-making processes, prioritizing cooperation across all levels of the organization (Deming, 2017). However, reducing traditional layers of supervision presents unique challenges, particularly in monitoring performance and motivating employees effectively. To address this problem, organizations often turn to tournament-based incentives, which promote a competitive, yet motivating environment that encourages high performance.

Nevertheless, the interplay between tournament and cooperation-based incentives intensifies the trade-off between individual and group benefits, potentially leading to an increase in free-riding behavior (Brandts and Riedl, 2020; Buser and Dreber, 2016). In other words, exposure to a competitive environment generates mindset spillovers, drawing attention to selfish norms and reducing cooperation. For organizations, this challenge reflects the need to develop and implement incentive structures and leadership models that effectively reconcile high-performing workers with the promotion of cooperation. Similarly, in educational settings, competition motivates students to strive for excellence, while the development of soft skills and collaborative abilities remains essential for maximizing their overall performance. Addressing the inherent trade-offs between individual and collective incentives is therefore essential for minimizing efficiency wastes.

In this study, we replicate two experimental conditions from Buser and Dreber (2016) to reproduce the negative impact of tournament incentives on cooperation. We further test the efficiency implication of a simple incentive mechanism that embeds prosocial elements into the competition. More specifically, we introduce donation programs within the competitive environment as a mechanism designed to counteract the negative spillovers of competition on prosocial behavior.

We design an online experiment in which participants are randomly assigned to one of four experimental conditions, each featuring a different incentive scheme for the completion of a simple real-effort task. In the *Control* condition, participants were compensated using a piece-rate payment scheme. In the remaining three conditions, compensation followed a winner-takes-all structure. In the *Competition* treatment, participants only received feedback on whether they won or lost the tournament. In the *Donation 10%* and *Donation 25%* treatments, winners of the competition were given the option to donate 10% or 25% of their prize to a charity, while losers observed the winner's decision. After the task, we measure their willingness to cooperate with a randomly selected group of participants with whom they had no prior interaction.

We find no evidence of detrimental effects of tournament incentives on prosocial behavior, questioning the existence of a negative spillover effect of competition on cooperation, as found in Buser and Dreber (2016). The latter result is supported by similar designs from the many-design study conducted by Huber et al. (2023)[1]. Furthermore, we find that donation programs within competitive environments do not produce any incremental effect on efficiency. Observing winners' prosocial behavior does not influence others' willingness to cooperate, leaving losers' attitudes unchanged across groups exposed to either positive or negative signals.

[1]Design IDs: HCA40, ICP06, IZU58, MUE79, NCW80, QLM89.

Our results contribute to the growing body of replication studies by re-examining the potential drawbacks of tournament incentives on cooperation. A concise yet influential body of literature highlights how competitive environments affect cooperation decisions. Generally, cooperation is negatively impacted when individuals have been previously exposed to tournament incentives. In a large online experiment, Buser and Dreber (2016) find that participants who compete in a tournament reduce subsequent contributions in a public goods game, even when matched with perfect strangers. When cooperation decisions involve former competitors, prosocial behavior is also reduced (Brandts and Riedl, 2020; Cason and Gangadharan, 2013; Chen, 2010; Kosse et al., 2025; Li et al., 2023; Moyal and Ritov, 2020). As a result, efficiency levels tend to be lower when cooperation decisions are made after a competitive effort. However, it is worth noting that Huber et al. (2023) examine competition effects across a wide range of moral (and prosocial) outcomes using many different experimental designs. In particular, the subset of designs that mimic our design yields mixed findings. This heterogeneity motivates the present replication exercise and underscores the need to further collect evidence from replications to better understand the mechanisms at play behind incentives' spillovers. In this respect, Chapter 1 of this dissertation provides a valid within-subject measure of the effect of competition on prosociality, further highlighting how results are sensitive with respect to the design and methodology adopted.

Several additional studies explore how different types of competition affect contributions in public goods games. For example, team incentives applied to inter-group competitions enhance efficiency from cooperation (Augenblick and Cunha, 2015; Bornstein and Gneezy, 2002; Markussen et al., 2014), and similarly, cooperation increases when individuals compete over better returns in public goods games (Angelovski et al., 2019; Bergantino et al., 2021; Colasante et al., 2019, 2017). However, when competitive and cooperative games are played simultaneously and share a common endowment, competition reduces overbidding in auctions but has no impact on cooperation decisions (Godoy et al., 2013; Savikhin and Sheremeta, 2013).

Our second contribution adds to research on organizational behavior by asking whether prosocial signals can promote cooperation alongside competitive incentives. Can the winners of a competition restore a cooperative mindset among losers by using donations to signal prosociality? Actions taken by tournament winners can send powerful signals to their previous competitors. In this regard, studies on leadership offer valuable insights by showing how discriminatory actions can hamper performance (Glover et al., 2017), or how setting good examples can foster cooperation, and vice versa (Güth et al., 2007; Johnson, 2015). Additionally, unfair signals or incentives can decrease motivation and productivity among team members, even if they are not personally affected by the injustice (Bandiera et al., 2005; Heinz et al., 2020). In this regard, winners' selection should not be based solely on performance but must also consider leadership qualities (Hoffman and Tadelis, 2021), strategic mindsets (Kaplan and Sorensen, 2017), and charisma (Antonakis et al., 2022).

Our focus, therefore, is on the behavioral changes that individuals undergo after exposure to a competitive environment. If agents become more selfish after a competition, we seek to restore individuals' mindset toward prosocial norms, reducing potential negative spillover effects.

Reduced cooperation hinders human and economic development by lowering efficiency. Embedding donation programs within a competitive environment offers the possibility to generate prosocial signals among competitors to promote altruistic norms. Additionally, donation programs are often used by firms to signal a commitment to a social cause to third parties (e.g., customers, partners, governments), and companies encourage their employees to participate in these programs. One example is given by corporate matching gifts, a type of donation scheme in which companies financially match donations made by their employees to nonprofit organizations. These incentives are intended to improve corporate social responsibility and employee engagement by generating prosocial signals.

In Section 2, we provide a description of the experimental design, Section 3 describes the data, Section 4 reports the results, and Section 5 concludes.

2.2 Experimental Design

Our design consists of four treatments, each sharing the same structure. In the first part, participants complete a two-minute real-effort task. The game is the same as that adopted by Buser and Dreber (2016), where participants are presented with a screen filled with sliders and asked to position them at the 50 mark before the time expires. Sliders vary in their value-length (see Figure A7 for details). Participants in the *Control* group are paid four cents for each slider they complete. Those in the *Competition* ("Feedback" condition in Buser and Dreber (2016)) treatment compete in groups of four members, where the highest performing subject is paid sixteen cents for each slider completed, while all other subjects receive zero. Participants receive feedback about the outcome of the contest. Together, the *Control* and *Competition* treatments constitute the direct replication of the conditions measuring the negative effect of tournament incentives on prosocial behavior. In two additional treatments (*Donation 10%* and *Donation 25%*), participants are compensated following the same incentive scheme of the *Competition* condition, but winners of the tournament can actively choose to donate a proportion of their prize to charity (American Red Cross). The winners' display shows the following message "You have won the competition. Do you want to donate X% (10% or 25%) of your prize to the American Red Cross?". The latter choice is revealed to losers in the feedback stage. The message that appears on their display states: "You have lost the competition. The winner chose (not) to donate X% of the prize to the American Red Cross". Importantly, participants are informed that winners' donation choices would be displayed to the matched loser (see instructions in Appendix A.2). We exogenously vary the proportion of the prize donation across the two *Donation* treatments (10% and 25%). We expect the shift of the donation percentage from 10% to 25% to have two effects: on one hand, because donating becomes more costly, the number of donations should decrease. On the other hand, increasing the amount of money donated increases the strength of the signal losers receive. In this regard, the results would inform about the effectiveness of donation programs and their potential heterogeneous effects on efficiency.

In the second stage, participants in all treatments are rematched in groups of four to play a one-shot public goods game. Importantly, the matching protocol follows a perfect stranger design, that is, no participants will be matched with someone they have interacted with in the task part. Each participant receives an initial allocation of 80 cents and has to make decisions regarding how much to keep for themselves and how much to contribute to the group. The money allocated to the group is doubled and evenly distributed among the four members of the group. The amount allocated to the group serves as our indicator of participants' willingness to cooperate. At the end of part II, participants fill in a brief survey indicating gender, risk attitude, and competitiveness. All instructions are provided at the beginning of the experiment, complying with the procedures adopted by Buser and Dreber (2016). The average earnings for participants are 1.25 USD for the public goods game and 1.29 USD for the slider task.

The sample size for our study was determined using a priori power analysis. In addition, we conducted simulations based on the empirical distribution of contributions to the public good reported by Buser and Dreber (2016), which produced the same sample size estimates (Campos-Mercade, 2024). To detect the same effect size as found in Buser and Dreber (2016) (Cohen's d = 0.29; piece rate vs. competition with feedback), with 80% statistical power and a significance level of 5%, a sample size of 370 participants (185 per group) is required. For our two *Donation* treatments, we hypothesize that the donation decisions will mitigate the existing spillovers, but only for winners and losers exposed to a pro-social signal (i.e., a donation by the winner)[2]. We compare each *Donation* treatment with the *Competition* condition by setting the minimum effect size we aim to detect at Cohen's d = 0.196, representing a 65% reduction in efficiency loss. To detect this effect size with 80% statistical power and a 5% significance level, a sample size of 812 participants is required (406 in the *Competition* condition and 406 winners/losers exposed to the prosocial signal). Since we expect approximately half of the tournament winners to donate, we doubled the sample size for the two *Donation* conditions. Consequently, the total planned recruitment target is 2,215 participants, distributed as follows: 185 in the *Control* condition, 406 in the *Competition* condition, and 812 in each of the two *Donation* conditions (10% and 25% donation rates).

We recruited participants on Prolific to take part in our experiment, programmed using oTree (Chen et al., 2016). Subjects voluntarily participated in our study. Randomization occurred at the session level to ensure enough participants played at the same time [3]. At the end of the experiment, one of the two stages is randomly selected for payment.

We pre-registered the following hypotheses using AsPredicted.org[4]. First, we aim at replicating the effect of competition on cooperation found in Buser and Dreber (2016). We expect participants to be less willing to cooperate when exposed to tournament incentives due to the presence of competitive primes that trigger selfish behavior (Buser and Dreber, 2016). Furthermore, we anticipate that prosocial signals introduced after the competition will mitigate the negative spillover effect of competition on cooperation. Specifically, if participants become more selfish after competing, we hypothesize that this effect can be counterbalanced by signals that encourage prosocial actions. Finally, we test whether the effect of donations on cooperation depends on the size of the donation, expecting that stronger signals lead to larger changes in subsequent contribution decisions.

[2]Importantly, we analyze winners and losers jointly, as significant differences between the two groups were not detected in the original Buser and Dreber (2016) paper.
[3]Information on session timings is provided in Table B1 of Appendix A.2.
[4]Pre-registration number AsPredicted #177381: https://aspredicted.org/zksx-x82v.pdf.

2.3 Data

Table 2.1 outlines the data and sample characteristics. The total number of participants recruited is 2,248, with 988 men and 1259 women. We collected twice as many observations in the *Donation* treatments compared to the *Competition* condition to ensure a sufficient number of participants exposed to the prosocial signal. Additionally, the similarities between the two *Donation* treatments allow us to further increase the number of observations by conducting a pooled analysis of donation decisions, regardless of their magnitude.

As in Buser and Dreber (2016), we found that women solved fewer sliders than men ($p<0.001$; t-test). Nevertheless, our sample performed on average worse than participants in the original study ($p<0.001$; t-test). In our setting, average public good contributions are higher for women ($p=0.038$; rank-sum test), but on average equal to that found in Buser and Dreber (2016) ($p=0.750$; rank-sum test). Our sample was on average older (36 vs. 30) than the original one ($p<0.001$; t-test). Lastly, in line with previous studies, men report higher scores for risk-seeking behavior ($p<0.001$; t-test).

During the experiment, some participants were excluded or dropped out. The total number of observations excluded from the analysis is 741, 533 of which failed to answer the comprehension checks of the public good game. The remaining 208 voluntarily left the study at different stages of the game after the treatment was revealed. We find that the latter do not represent a concern in terms of selective attrition ($p=0.291$; Fisher's exact test).

Finally, we provide information regarding the minimum detectable effect in the context of the replication exercise. With 197 and 442 participants in the *Control* and *Competition* conditions, respectively, and a pooled standard deviation of 34.2 (consistent with the original study), the minimum detectable effect corresponds to a Cohen's d of 0.24, assuming 80% statistical power and a 5% significance level. This represents 83% of the effect size reported in the original study.

The outline of the results is the following. First, we describe the result of the replication exercise, and second, we report the results of the two additional treatments. We then comment on differences in performance, and finally, we study gender differences in contributing behavior.

	Sample	Men	Women
Observations			
Control	197	89	108
Competition	442	184	258
Donation 10%	798	337	460
Donation 25%	811	378	433
Total	2247	988	1259
Variable means			
Performance	24	26	23
Contribution	45	43	46
Age	36	35	37
Risk Seeking	5	6	5

Table 2.1: Descriptive Statistics

2.4 Results

In Figure 2.1[5], we present the average contributions to the public good reported by Buser and Dreber (2016) and our own data. Participants in the original study contribute an average of \$0.52 in the *Control* condition and \$0.42 in the *Competition* treatment (referred to as the "Feedback" condition in the original paper). Their observed differences are statistically significant ($p < 0.001$; rank-sum test). In comparison to the original data, participants in our sample contribute on average \$0.04 less in the *Control* condition (averaging \$0.48) and \$0.05 more in the *Competition* group (averaging \$0.47). In our data, we find no significant difference between the *Control* and *Competition* treatments ($p = 0.858$).

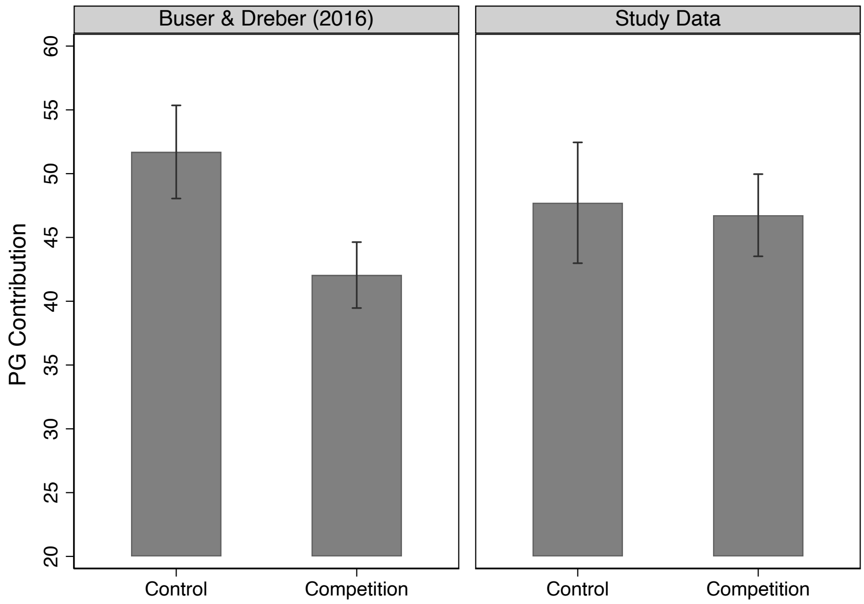


Figure 2.1: Public Good Game Contributions: Replication Exercise

In Figure 2.2, we present the differences between winners and losers in the *Competition* treatment (not preregistered[6]). In both studies, no significant differences are detected between the two groups. In the study conducted by Buser and Dreber (2016), winners contribute an average of \$0.44, whereas losers contribute \$0.42. In our study, both groups exhibit higher average contributions: winners contribute \$0.50, while losers contribute \$0.46. Although winners contribute slightly more than losers, the differences are not statistically significant ($p = 0.360$).

In the following analyses, we present the results of the two *Donation* treatments. Each *Donation* condition consists of two main groups: a prosocial group, comprising winners who made a donation and losers who received the prosocial signal, and a non-prosocial group, comprising winners who did not make a donation and losers who received the non-prosocial signal. Figure 2.3 illustrates the average contributions to the public good game across all treatments, distinguishing between prosocial and non-prosocial groups. In each group, contributions are lower on

[5] All figures presented henceforth include error bars representing 95% confidence intervals. "PG" stands for public good.
[6] The comparison between losers and winners was not part of the original study plan, but it was later added as it provided valuable context for interpreting our results.

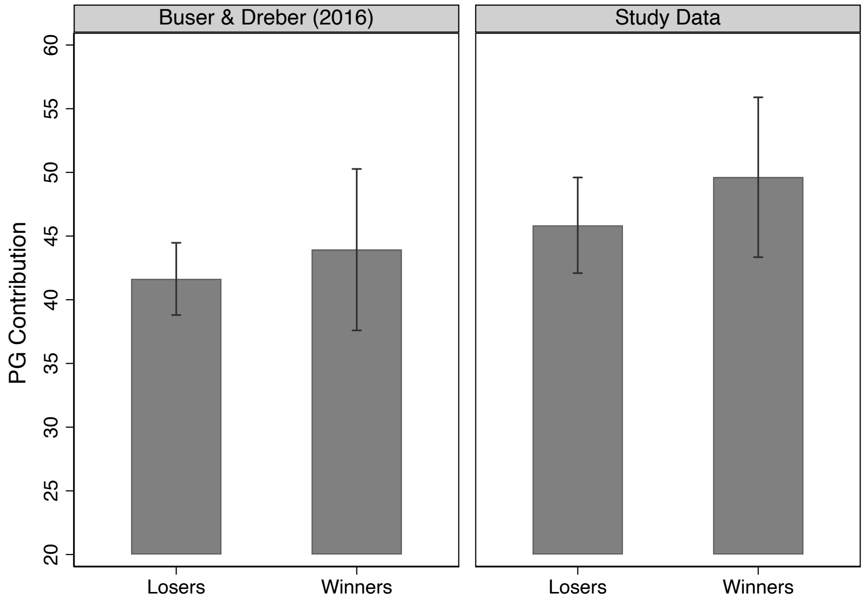


Figure 2.2: Public Good Game Contributions: Replication Winners and Losers Differences

average only for individuals (winners and losers) included in the non-prosocial condition.

In our setting, the overall difference in contributions across treatments is statistically significant ($p = 0.002$; Kruskal–Wallis test). Specifically, in the *Donation* treatments, participants in the prosocial group contribute an average of \$0.49 and \$0.50 in the *Donation* 10% and 25% treatments, respectively. However, these differences are not statistically significant when compared to the control group ($p = 0.592$ and $p = 0.430$; rank-sum test). Conversely, participants in the non-prosocial group contribute an average of \$0.41 (*Donation 10%*) and \$0.42 (*Donation 25%*), which is significantly lower than the *Control* group's average contribution of \$0.48 ($p = 0.040$ and $p = 0.074$). Furthermore, these contributions are also significantly lower than those of participants in the prosocial group ($p = 0.004$ and $p = 0.009$).

In Table 2.2, we summarize the previous results using ordinary least squares (OLS) regressions[7]. Column (1) presents the baseline regression outcomes using treatment dummies, with the *Control* condition as the reference category. Column (2) introduces controls for gender, performance, and age, while Column (3) further incorporates a control for risk attitudes. The findings do not support the hypothesis that prosocial signals mitigate the (statistically non-significant) efficiency gap introduced by competitive settings, nor do they indicate that contribution levels are significantly higher compared to the *Competition* treatment. This outcome remains consistent across both the 10% and 25% *Donation* treatments. Conversely, non-prosocial signals, observed in both *Donation* conditions, result in significantly lower contributions by participants relative to all other treatment conditions, with the exception of comparisons between the two non-prosocial treatments themselves. Compared to the *Control*, *Competition*, and both proso-

[7]Consistent with the pre-registration, we estimate the main specifications following the empirical approach of Buser and Dreber (2016), using robust standard errors. To account for potential interdependencies among participants who shared the same competitive experience in the first part, we additionally re-estimate the regressions clustering standard errors at the competition-group level. The results remain unchanged in magnitude and statistical significance. See Appendix B.2 for details.

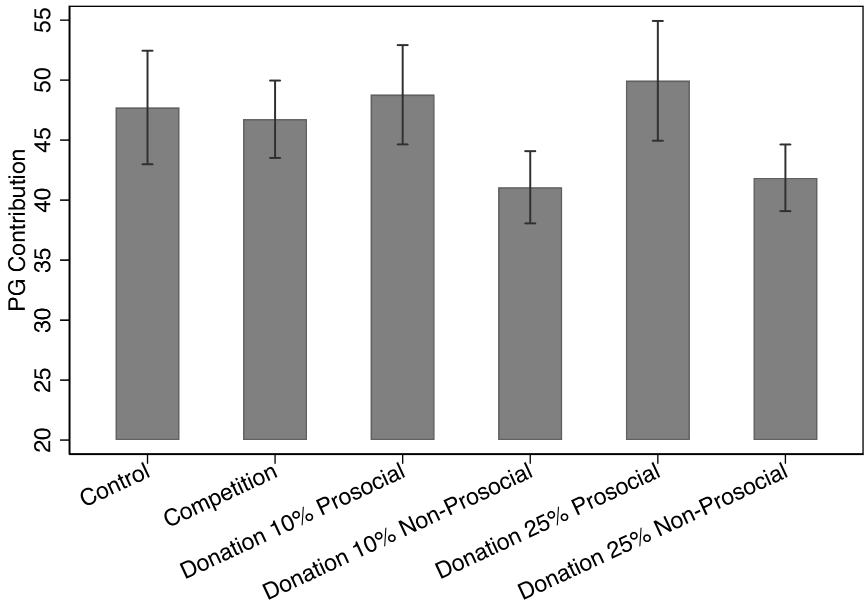


Figure 2.3: Average Contributions by Treatment

cial *Donation* conditions, participants in the non-prosocial group contribute 16.72% less in the *Donation 10%* condition and 14.53% less in the *Donation 25%* condition. This treatment effect persists even when controlling for observables, including gender, performance, age, and self-reported risk-seeking behavior. Men participants are found to contribute less than women, while higher performance, age, and greater self-reported risk-seeking tendencies are associated with higher contribution levels.

In summary, while all observed effect signs align with our expectations, only participants in the non-prosocial group lead to a measurable decline in cooperation. Notably, this effect originates from the compounded reactions of losers who receive the signal and winners who make the donation decision. Therefore, we split our sample between losers and winners to disentangle the exogenous effect of the signal received for losers and the endogenous impact of making a donation decision for winners (not preregistered).

Figure 2.4 illustrates the differences in contributions between winners and losers across treatment conditions. Although splitting the sample groups reduces statistical power, two results emerge clearly. First, competition winners display differential contributions based on whether they sent a prosocial or non-prosocial signal. Interestingly, the magnitude of the donation (10% versus 25%) appears negligible, indicating that contributions do not linearly increase with the size of the donation. Second, the received signal has not significant behavioral impacts on losers, suggesting that observing winners' prosocial actions does not facilitate higher efficiency levels when tournament and cooperation incentives interact with each other.

The previously observed reduction in contributions among participants in the non-prosocial group can therefore be attributed to specific underlying factors. As previously discussed, the compounded reactions of both losers and winners likely contribute to observe lower contributions in this group, but this effect is primarily driven by winners who are endogenously selected as low

	(1) Contribution	(2) Contribution	(3) Contribution
Competition	-0.974	-1.237	0.230
	(2.905)	(2.909)	(2.863)
Donation 10% Prosocial	1.063	1.051	1.817
	(3.190)	(3.186)	(3.159)
Donation 10% Non-Prosocial	-6.650**	-6.668**	-6.018**
	(2.847)	(2.852)	(2.809)
Donation 25% Prosocial	2.227	2.083	2.683
	(3.483)	(3.465)	(3.425)
Donation 25% Non-Prosocial	-5.863**	-5.782**	-4.967*
	(2.785)	(2.794)	(2.754)
Male		-3.443**	-5.769***
		(1.505)	(1.521)
Performance		0.336***	0.412***
		(0.087)	(0.087)
Age		0.110**	0.117**
		(0.047)	(0.048)
Risk			2.123***
			(0.285)
Constant	47.716***	37.109***	23.598***
	(2.399)	(3.798)	(4.158)
Observations	2247	2246	2246
R-squared	0.009	0.017	0.042

Standard errors in parentheses

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

Table 2.2: OLS Regression: Contributions in the Public Good Game

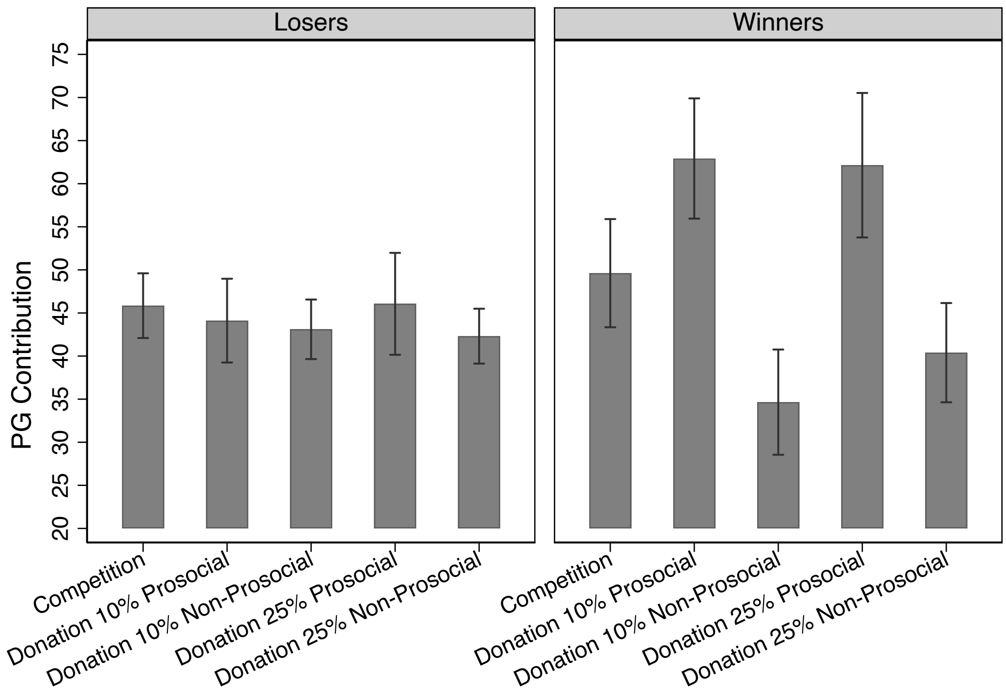


Figure 2.4: Average Contributions by Treatment: Winners and Losers Differences

or high contributors. Yet, Table 2.2 shows a stronger negative effect in the non-prosocial group, but not an equally positive one in the prosocial group. For this reason, selection might not be the only mechanism at play, suggesting that the act of making a donation decision (regardless of whether the signal is prosocial) may itself have an effect. We report non-preregistered results to investigate winners' behavior. Yet, due to the limited number of winners in the sample, statistical inference regarding alternative mechanisms remains speculative.

To explore other potential mechanisms, we test for differences among winners across the *Competition* and *Donation* treatments, excluding the selection channel by pooling winners who chose to donate with those who did not. Additionally, we combine winners from both *Donation* treatments to increase statistical power, as no significant differences are observed between the 10% and 25% *Donation* treatments. In Table 2.3, we regress contributions in the public goods game on a single treatment variable (*Donation*) within the winner sub-sample, using the *Competition* treatment as the reference category. This approach allows us to estimate the effect of making a donation decision in a competitive setting. If the contribution gap observed in Figure 2.4 were solely a result of selection, we would not expect any difference between the two types of winners. To some extent, this holds true: regressions in columns (1) and (2) do not report any statistically significant differences between winners in the *Donation* treatment and winners in the *Competition* treatment. However, when controlling for risk attitudes, there is some indication that winners in the *Donation* treatment contribute, on average, less than those in the *Competition* treatment. This finding confirms that selection is the main mechanism at play, but it may not be the only one driving the differences in contributions between winners who donate and those who do not.

In summary, we provide indicative evidence that winners who do not donate decrease their contributions to a greater extent than the increase observed among winners who donate. One plausible mechanism consists of a mixture of self-signaling (Gneezy, 2023) and negativity bias, which may alter individuals' self-perception. In this context, not donating could signal to winners that they are not prosocial types, potentially reducing their willingness to contribute in the public goods game more than a counterfactual scenario without any donation decision. On the other hand, among the mechanisms explaining the selection into types of contributors, we may consider a "taste for consistency" between the two winners' choices, which could lead participants to act prosocially if they donate and vice versa (Falk and Zimmermann, 2013).

Another interesting advantage of donation programs is their potential spillover effect on workers' engagement. If workers are more motivated by donation initiatives, we should observe an increase in the number of sliders completed. However, in line with Buser and Dreber (2016), performance in the slider task (Figure 2.5) does not vary across groups, suggesting that our treatments have no effect on performance, neither due to incentives stemming from competition nor from the implemented donation program (p=0.4554, Kruskal–Wallis test). In our case, the task length (two minutes) may impede the identification of variations in performance, leaving no room to properly test for the differential effect on efficiency stemming from the adoption of tournament incentives.

Competition also influences behavior across genders. Previous research has examined how

	(1) Contribution	(2) Contribution	(3) Contribution
Donation	-5.006	-4.994	-5.745*
	(3.631)	(3.603)	(3.454)
Male		-1.769	-4.643
		(3.392)	(3.411)
Performance		0.530**	0.593**
		(0.260)	(0.258)
Age		0.075	0.115
		(0.173)	(0.167)
Risk			2.701***
			(0.627)
Observations	493	493	493
R-squared	0.003	0.012	0.049

Standard errors in parentheses

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

Table 2.3: OLS Regression: Test for Selection

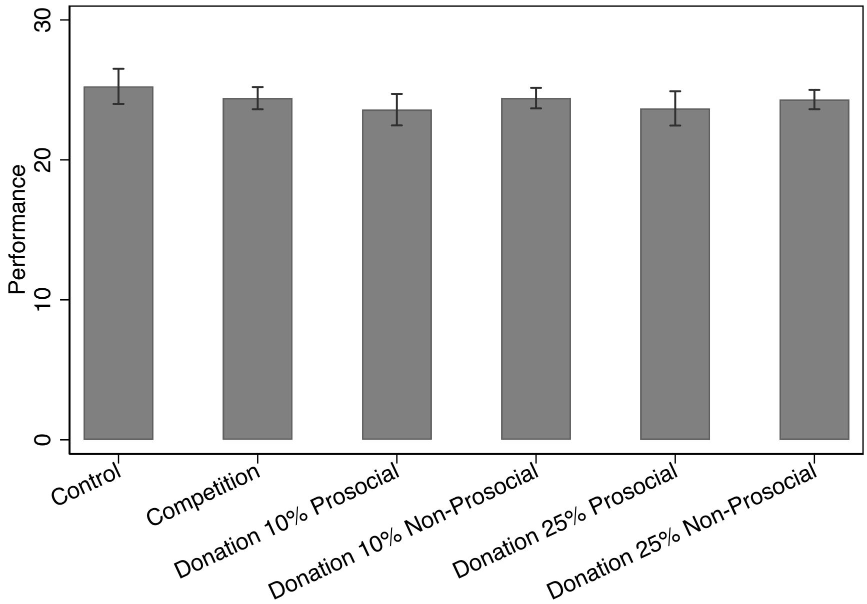


Figure 2.5: Performance in the Slider Task

men and women respond to incentives in competitive versus non-competitive environments, consistently showing that men tend to select in tournaments more often than women (Croson and Gneezy, 2009; Niederle and Vesterlund, 2007). Hence, by imposing tournament incentives, individuals' differing preferences for competition may interact with the competitive environment, disproportionately affecting those participants who would have opted out if given a choice. As indicated in the pre-registration, we also replicate gender differences in contributing behavior due to previous exposure to competition. In our study, we do not observe any significant gender differences in contributing behavior when comparing the *Control* and the *Competition* condition. Similarly, Buser and Dreber (2016) identify a difference in behavior only among competition losers, but when winners and losers are analyzed together, such differences disappear. Considering our additional treatment conditions, contributions vary significantly among women (p = 0.014; Kruskal-Wallis test) and among men (p = 0.038). However, the endogenous selection into prosocial and non-prosocial contributors reasonably explains most of the gap observed, once again negating the existence of meaningful differences. Using Mann-Whitney tests comparing contributions in each treatment with those in the *Control* condition, we find no significant differences within either gender, as summarized by the p-values reported in Table 2.4.

Treatment	Control (Female)	Control (Male)
Competition	0.404	0.549
Donation 10% Prosocial	0.825	0.598
Donation 10% Non-Prosocial	0.066	0.279
Donation 25% Prosocial	0.534	0.667
Donation 25% Non-Prosocial	0.153	0.315

Table 2.4: Treatment Effects - Gender

Figure 2.6 displays the contribution levels for men and women separately, along with the reference results from the Buser and Dreber (2016) analysis. One difference between our data and the one collected by Buser and Dreber (2016) emerges clearly: men in the *Control* condition contribute less than men in the original study. To a lesser extent, the opposite tendency can be observed in the *Competition* treatment, leading to the results of the unsuccessful replication. Finally, to formalize the results on gender differences, Table 2.5 reports the regression results (pooling the two *Donation* treatments) by gender. Although only female participants in the non-prosocial signal group significantly decreased their contributions, none of the gender-based differences in contributions achieved statistical significance ($p > 0.05$; Wald tests on gender–treatment interaction).

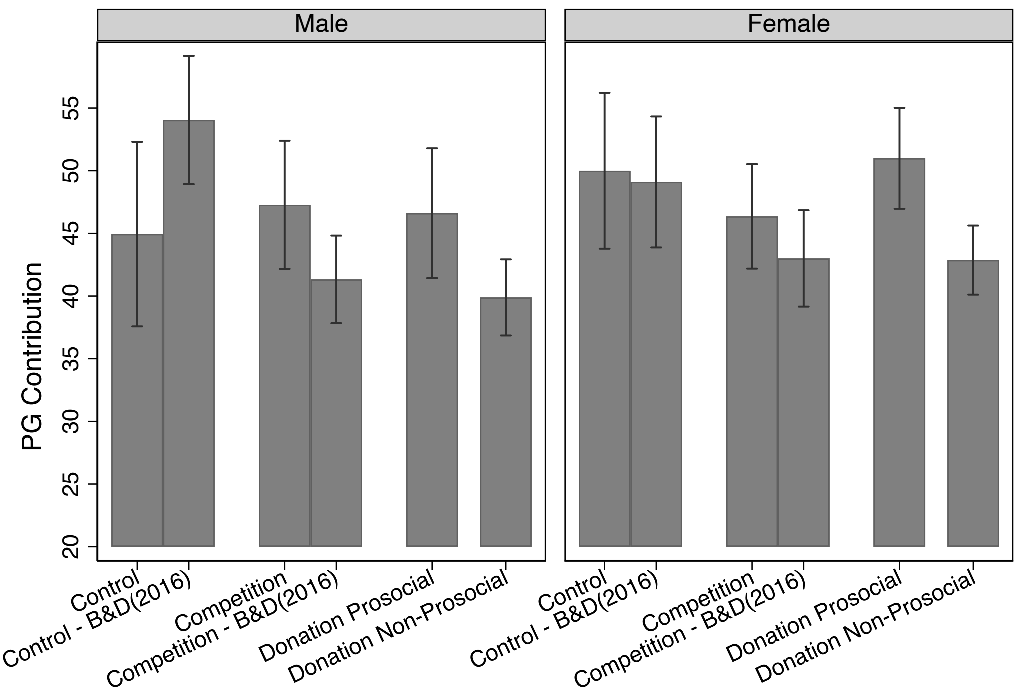


Figure 2.6: Public Good Game contributions (by gender)

	(1) Contribution (Men)	(2) Contribution (Women)	(3) Contribution (Men)	(4) Contribution (Women)	(5) Contribution (Men)	(6) Contribution (Women)
Competition	2.339	-3.643	2.688	-4.173	4.205	-2.780
	(4.511)	(3.777)	(4.542)	(3.774)	(4.531)	(3.658)
Donation (Prosocial)	1.935	0.992	1.804	0.926	2.768	1.334
	(4.530)	(3.737)	(4.564)	(3.736)	(4.593)	(3.616)
Donation (Non-Prosocial)	-5.055	-7.132**	-4.900	-7.255**	-4.280	-6.418*
	(4.005)	(3.429)	(4.067)	(3.422)	(4.083)	(3.295)
Performance			0.369***	0.298**	0.445***	0.365***
			(0.118)	(0.132)	(0.119)	(0.131)
Age			0.083	0.122**	0.104	0.120**
			(0.100)	(0.052)	(0.098)	(0.053)
Risk					1.800***	2.387***
					(0.446)	(0.372)
Constant	44.944***	50.000***	32.292***	38.788***	17.889**	24.477***
	(3.694)	(3.129)	(6.319)	(5.045)	(7.291)	(5.335)
Observations	987	1259	987	1259	987	1259
R^2	0.009	0.010	0.019	0.015	0.036	0.047

Standard errors in parentheses

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

Table 2.5: OLS - Gender

2.5 Conclusion

In this study, we investigate whether donation programs can mitigate the previously documented negative spillover effects of tournament incentives on cooperative behavior. To answer this question, we replicate the piece rate and competition conditions (with feedback) from the experiment conducted by Buser and Dreber (2016) and introduce two additional treatments that incorporate donation programs to examine whether charity initiatives can mitigate the adverse spillover effects of competition on cooperation. The findings from our replication exercise indicate that competition does not significantly alter individuals' attitudes toward cooperation. With respect to the Buser and Dreber (2016) study, our results can be partially attributed to male participants, who exhibit lower contributions in the piece-rate condition and higher contributions under competition. Our results are further supported by a recent meta-analysis, which studies the impact of competition on morals (Huber et al., 2023), ultimately suggesting that the effect of tournament incentives in subsequent behavior requires further exploration. Taken together with the results of Chapter 1, the present findings allow us to draw a more nuanced picture of competitive spillovers. Huber et al. (2023) cover a wide range of experimental designs and moral outcomes; not all of their designs target prosociality, and the cooperation-related evidence within that meta-analysis is itself heterogeneous. Our inability to replicate the negative spillover documented by Buser and Dreber (2016) is consistent with this broader uncertainty. While Chapter 1 does report a significant negative spillover, the latter is measured in a qualitatively different setting. The contrast between the two chapters points to the competitive mechanism as a critical moderator. Do tournament incentives lead agents to be less prosocial? The answer to this question now requires knowing what the underlying mechanisms are across different treatments. A clear answer can be provided by jointly observing the effect of incentives and the mechanisms driving it, providing clear information about what type of competitive settings actually generate a loss of social surplus from cooperation. In this respect, while Huber et al. (2023) find that competition does not impact subsequent cooperation, Kosse et al. (2025) detect a negative spillover of competitive environment among students in the field.

In two additional treatments, we disentangle the role of signals by examining whether donations made by competition winners can increase cooperation. Our results reveal no substantial effect of prosocial signals. In our sample, reduced contributions are observed more frequently within the subgroup of participants who engage in non-prosocial interactions, both as senders (winners) and receivers (losers) of the signal. Nonetheless, additional analysis of differences between winners and losers in cooperation decisions suggests that the treatment effects are primarily driven by winners, who are endogenously self-selected into prosocial and non-prosocial groups. This pattern implies that mechanisms such as moral licensing and other competition effects do not significantly influence winners' donation decision and their subsequent cooperation. Therefore, we conclude that the observed variation in contributions among subjects in the non-prosocial group mainly results from the self-selection of winners into contributor types, and it is only reinforced by losers' reactions to non-prosocial signals.

The findings from this study have implications for designing workplace and organizational policies that foster cooperation in competitive environments. While competition is often used as a motivational tool, it is commonly thought to suppress cooperative behavior. However, our inability to replicate the existence of negative spillovers suggests that such effects may not be as robust or generalizable as expected. Policies that incorporate prosocial elements, such as donation programs, although not shown here to significantly impact cooperative behavior, could still play a role in shaping an organizational culture that values both individual success and collective goals, but does not interact with tournament incentives. This suggests that organizations may not need to overly worry about competition undermining cooperation, especially if they create environments that naturally encourage both teamwork and individual achievement.

The central question for future research on detecting and quantifying spillovers of tournament incentives is why we observe effects in some contexts but not in others. Should we be concerned at all when using tournament incentives? If so, which financial and social incentives best sustain motivation without negatively influencing subsequent decisions? Two issues stand out in the current literature. First, while the nature of incentives clearly shapes immediate decisions, it remains unclear how these incentives operate in subsequent choices. Second, when observing competitive environments, future studies should account for self-selection into competition to better control for individual preferences, since tournament incentives can motivate effectively only a specific group of the population.

References

Angelovski, A., Neugebauer, T., and Servátka, M. (2019). Rank-order competition in the voluntary provision of impure public goods. *Economic Inquiry*, 57(4):2163–2183.

Antonakis, J., d'Adda, G., Weber, R. A., and Zehnder, C. (2022). "just words? just speeches?" on the economic value of charismatic leadership. *Management Science*, 68(9):6355–6381.

Augenblick, N. and Cunha, J. M. (2015). Competition and cooperation in a public goods game: A field experiment. *Economic inquiry*, 53(1):574–588.

Bandiera, O., Barankay, I., and Rasul, I. (2005). Social preferences and the response to incentives: Evidence from personnel data. *The Quarterly Journal of Economics*, 120(3):917–962.

Bergantino, A. S., Gil-Gallen, S., and Morone, A. (2021). Do risk and competition trigger conditional cooperation? evidence from public good experiments. *Annals of Public and Cooperative Economics*.

Bornstein, G. and Gneezy, U. (2002). Price competition between teams. *Experimental Economics*, 5:29–38.

Brandts, J. and Riedl, A. (2020). Market interaction and efficient cooperation. *European Economic Review*, 121:103318.

Buser, T. and Dreber, A. (2016). The flipside of comparative payment schemes. *Management Science*, 62(9):2626–2638.

Campos-Mercade, P. (2024). Power analysis through simulations in stata: A step-by-step guide. Technical report, Working Paper.

Cason, T. N. and Gangadharan, L. (2013). Cooperation spillovers and price competition in experimental markets. *Economic Inquiry*, 51(3):1715–1730.

Chen, D. L. (2010). Markets and morality: How does competition affect moral judgment. Technical report, Working Paper, Duke University School of Law.

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

Colasante, A., García-Gallego, A., Georgantzis, N., Morone, A., and Temerario, T. (2019). Intragroup competition in public good games: The role of relative performance incentives and risk attitudes. *Journal of Public Economic Theory*, 21(5):847–865.

Colasante, A., García-Gallego, A., Morone, A., and Temerario, T. (2017). The utopia of cooperation: does intra-group competition drive out free riding. *Working Papers*, 8.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic literature*, 47(2):448–474.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The quarterly journal of economics*, 132(4):1593–1640.

Falk, A. and Zimmermann, F. (2013). A taste for consistency and survey response behavior. *CESifo Economic Studies*, 59(1):181–193.

Glover, D., Pallais, A., and Pariente, W. (2017). Discrimination as a self-fulfilling prophecy: Evidence from french grocery stores. *The Quarterly Journal of Economics*, 132(3):1219–1260.

Gneezy, U. (2023). *Mixed signals: how incentives really work*. Yale University Press.

Godoy, S., Morales, A. J., and Rodero, J. (2013). Competition lessens competition: An experimental investigation of simultaneous participation in a public good game and a lottery contest game with shared endowment. *Economics Letters*, 120(3):419–423.

Güth, W., Levati, M. V., Sutter, M., and Van Der Heijden, E. (2007). Leading by example with and without exclusion power in voluntary contribution experiments. *Journal of Public Economics*, 91(5-6):1023–1042.

Heinz, M., Jeworrek, S., Mertins, V., Schumacher, H., and Sutter, M. (2020). Measuring the indirect effects of adverse employer behaviour on worker productivity: A field experiment. *The Economic Journal*, 130(632):2546–2568.

Hoffman, M. and Tadelis, S. (2021). People management skills, employee attrition, and manager rewards: An empirical analysis. *Journal of political economy*, 129(1):243–285.

Huber, C., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Weitzel, U., Abellán, M., Adayeva, X., Ay, F. C., Barron, K., et al. (2023). Competition and moral behavior: A meta-analysis of forty-five crowd-sourced experimental designs. *Proceedings of the National Academy of Sciences*, 120(23):e2215572120.

Johnson, R. R. (2015). Leading by example: Supervisor modeling and officer-initiated activities. *Police Quarterly*, 18(3):223–243.

Kaplan, S. N. and Sorensen, M. (2017). Are ceos different? characteristics of top managers. Technical report, National Bureau of Economic Research.

Kosse, F., Rajan, R., and Tincani, M. (2025). The persistent effect of competition on prosociality. *Journal of the European Economic Association*.

Li, Y., Liu, Z., Wang, Y., Derrington, E., Moisan, F., and Dreher, J.-C. (2023). Spillover effects of competition outcome on future risky cooperation. *Scientific Reports*, 13(1):5535.

Markussen, T., Reuben, E., and Tyran, J.-R. (2014). Competition, cooperation and collective choice. *The Economic Journal*, 124(574):F163–F195.

Moyal, A. and Ritov, I. (2020). The effect of contest participation and contest outcome on subsequent prosocial behavior. *Plos one*, 15(11):e0240712.

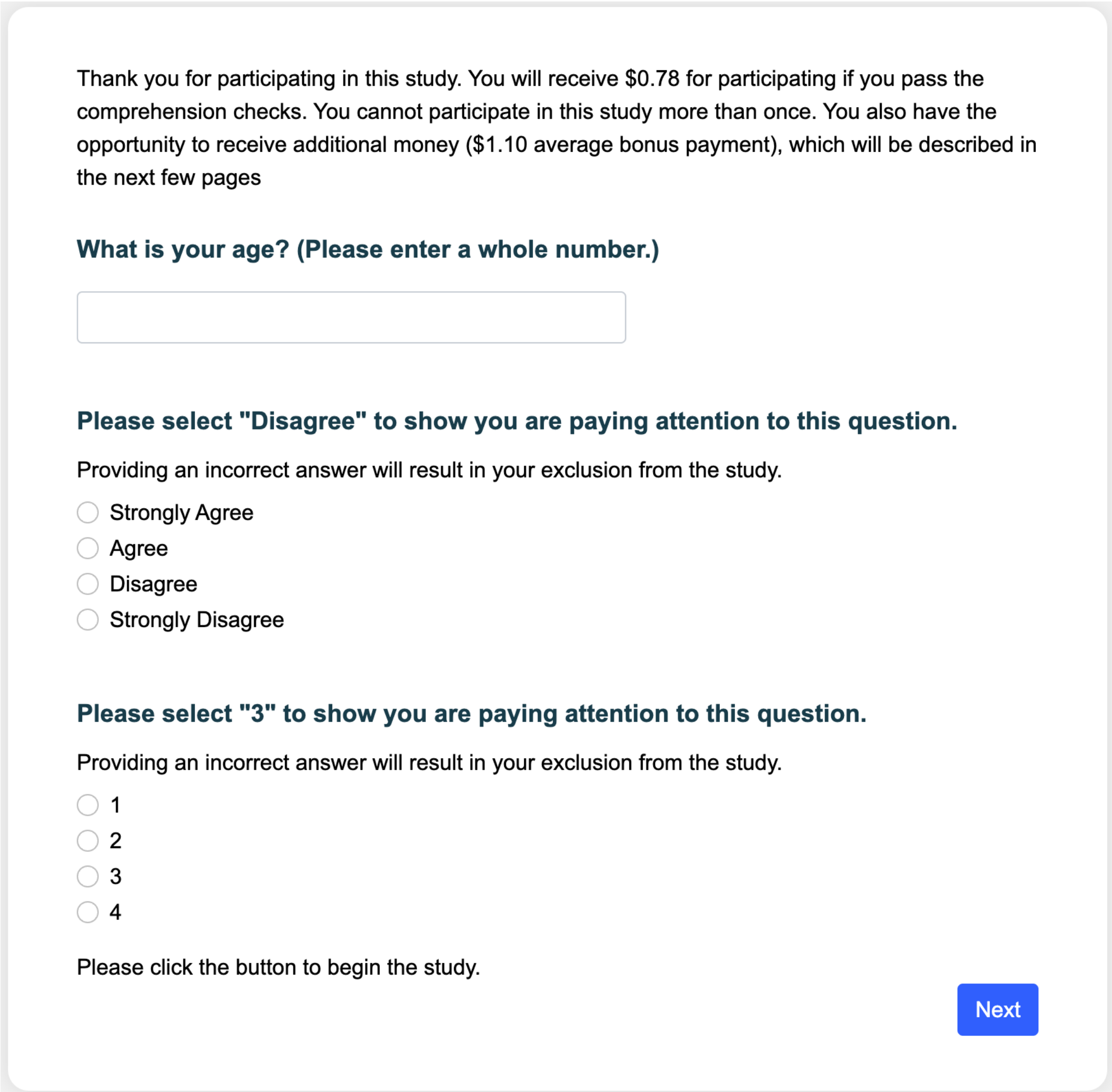
Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The quarterly journal of economics*, 122(3):1067–1101.

Savikhin, A. C. and Sheremeta, R. M. (2013). Simultaneous decision-making in competitive and cooperative environments. *Economic Inquiry*, 51(2):1311–1323.

Appendix A

Here we report screenshots and instructions from the experiment conducted.

A.1 Screenshots



Thank you for participating in this study. You will receive $0.78 for participating if you pass the comprehension checks. You cannot participate in this study more than once. You also have the opportunity to receive additional money ($1.10 average bonus payment), which will be described in the next few pages

What is your age? (Please enter a whole number.)

Please select "Disagree" to show you are paying attention to this question.

Providing an incorrect answer will result in your exclusion from the study.

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree

Please select "3" to show you are paying attention to this question.

Providing an incorrect answer will result in your exclusion from the study.

- 1
- 2
- 3
- 4

Please click the button to begin the study.

Next

Figure A1: Welcome and Attention Checks

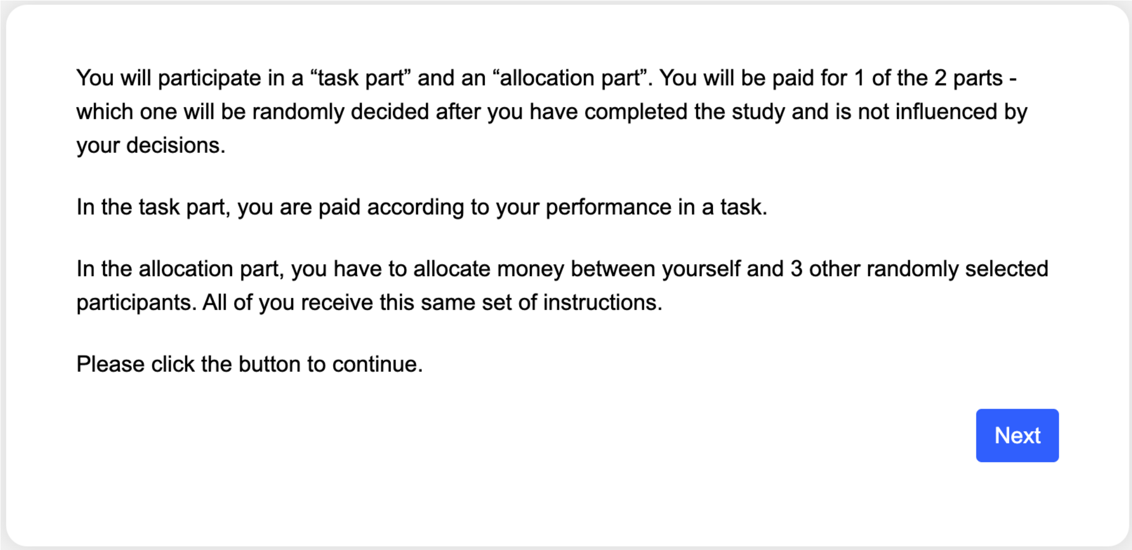


Figure A2: Experiment General Description

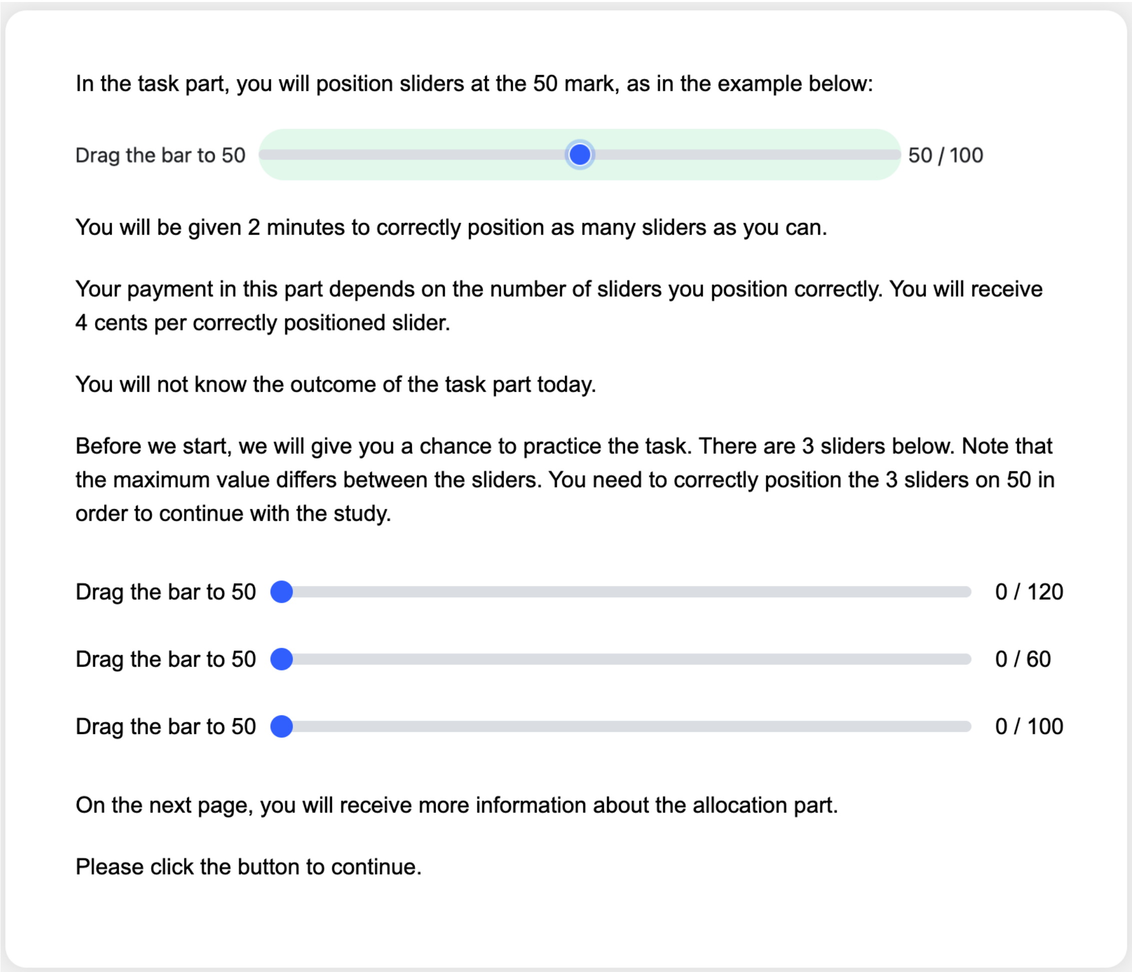
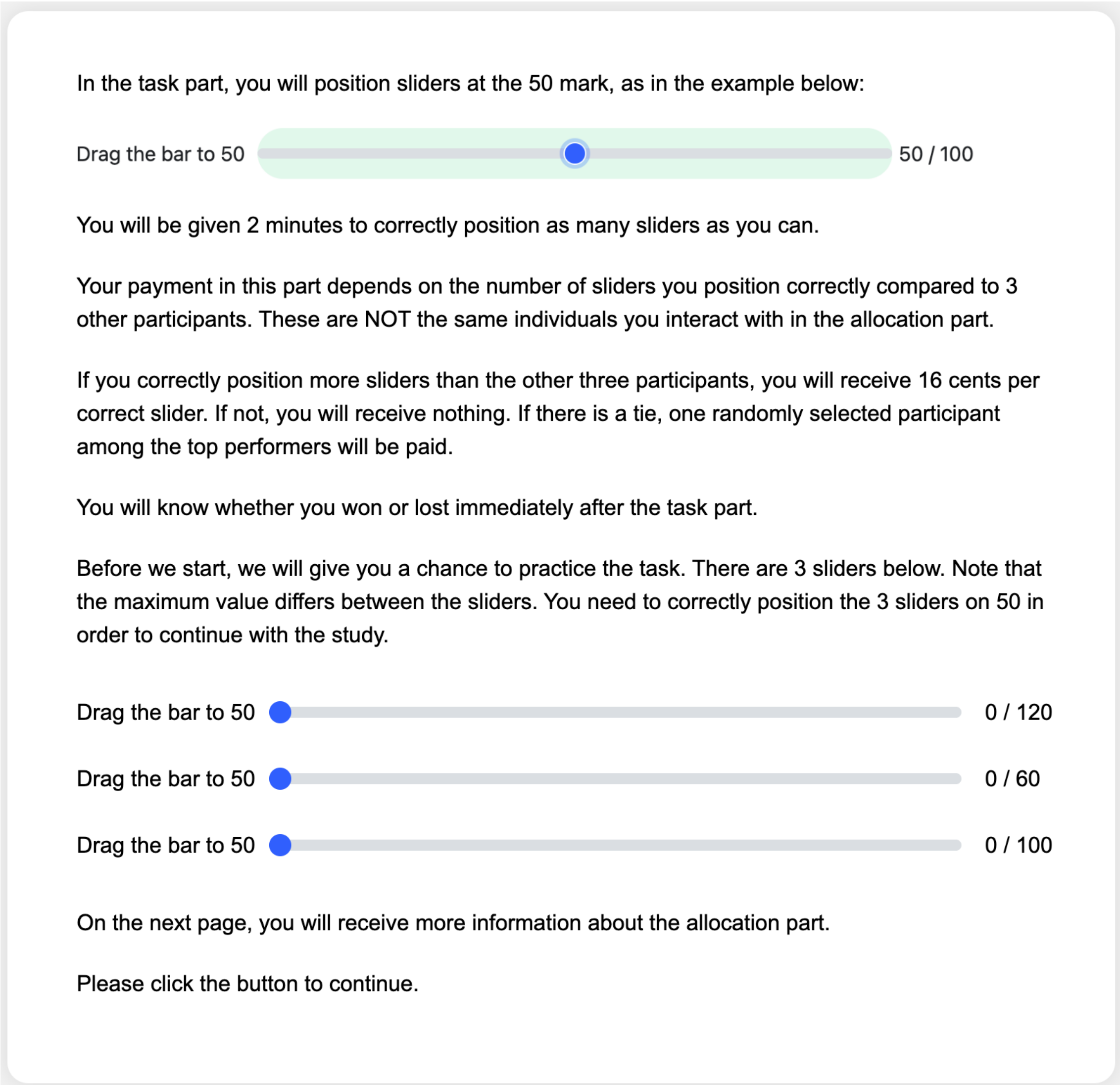


Figure A3: Slider Task Instructions: Piece Rate


In the task part, you will position sliders at the 50 mark, as in the example below:

Drag the bar to 50 50 / 100

You will be given 2 minutes to correctly position as many sliders as you can.

Your payment in this part depends on the number of sliders you position correctly compared to 3 other participants. These are NOT the same individuals you interact with in the allocation part.

If you correctly position more sliders than the other three participants, you will receive 16 cents per correct slider. If not, you will receive nothing. If there is a tie, one randomly selected participant among the top performers will be paid.

You will know whether you won or lost immediately after the task part.

Before we start, we will give you a chance to practice the task. There are 3 sliders below. Note that the maximum value differs between the sliders. You need to correctly position the 3 sliders on 50 in order to continue with the study.

Drag the bar to 50 0 / 120

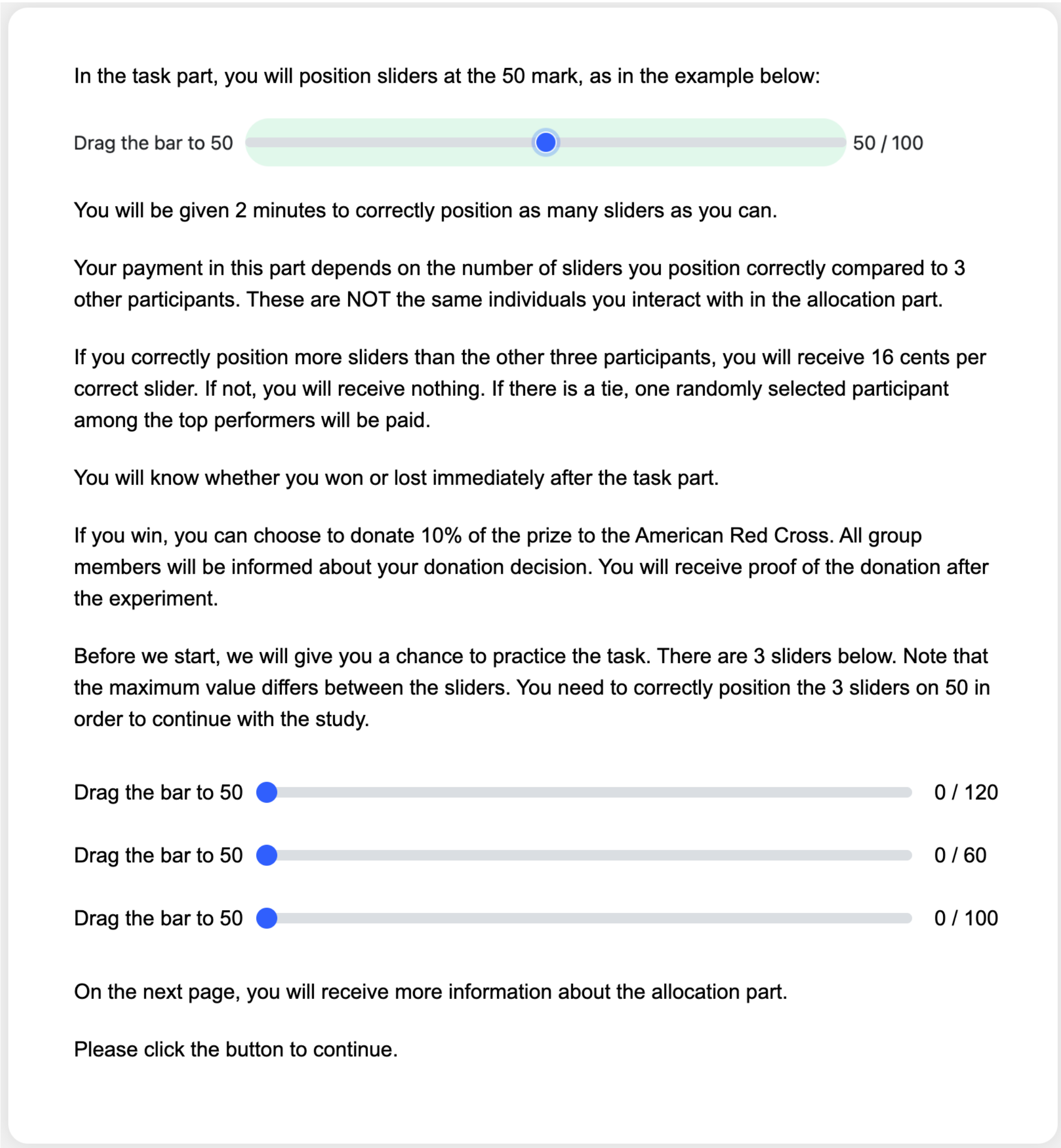
Drag the bar to 50 0 / 60

Drag the bar to 50 0 / 100

On the next page, you will receive more information about the allocation part.

Please click the button to continue.

Figure A4: Slider Task Instructions: Competition



In the task part, you will position sliders at the 50 mark, as in the example below:

Drag the bar to 50 50 / 100

You will be given 2 minutes to correctly position as many sliders as you can.

Your payment in this part depends on the number of sliders you position correctly compared to 3 other participants. These are NOT the same individuals you interact with in the allocation part.

If you correctly position more sliders than the other three participants, you will receive 16 cents per correct slider. If not, you will receive nothing. If there is a tie, one randomly selected participant among the top performers will be paid.

You will know whether you won or lost immediately after the task part.

If you win, you can choose to donate 10% of the prize to the American Red Cross. All group members will be informed about your donation decision. You will receive proof of the donation after the experiment.

Before we start, we will give you a chance to practice the task. There are 3 sliders below. Note that the maximum value differs between the sliders. You need to correctly position the 3 sliders on 50 in order to continue with the study.

Drag the bar to 50 0 / 120

Drag the bar to 50 0 / 60

Drag the bar to 50 0 / 100

On the next page, you will receive more information about the allocation part.

Please click the button to continue.

Figure A5: Slider Task Instructions: Competition + Donation

In the allocation part, each person in your group with 3 other people is given 80 cents (in addition to the 60 cents you received already for participating).

You each decide how much of your 80 cents to keep for yourself, and how much (if any) to contribute to the group's common project (in increments of 10 cents: 10, 20, 30, etc).

All money contributed to the common project is doubled, and then split evenly among the 4 group members.

Thus, for every 2 cents contributed to the common project, each group member receives 1 cent.

If everyone contributes all of their 80 cents, everyone's money will double: each of you will earn 160 cents.

But if everyone else contributes their 80 cents, while you keep your 80 cents, you will earn 200 cents, while the others will earn only 120 cents. That is because for every 2 cents you contribute, you get only 1 cent back. Thus you personally lose money on contributing.

The other people are REAL and will really make a decision - there is no deception in this study.

Once you and the other people have chosen how much to contribute, the allocation part is over.

Please, answer the following questions. Providing an incorrect answer twice will result in your exclusion from the study. After this, please click the button to start with the task part..

Once the sliders appear the 2 minutes for the task part will start.

What level of contribution earns the highest payoff for the group as a whole?

○ 0 ○ 10 ○ 20 ○ 30 ○ 40 ○ 50 ○ 60 ○ 70 ○ 80

What level of contribution earns the highest payoff for you personally?

○ 0 ○ 10 ○ 20 ○ 30 ○ 40 ○ 50 ○ 60 ○ 70 ○ 80

Next

Figure A6: Public Good Game Instructions

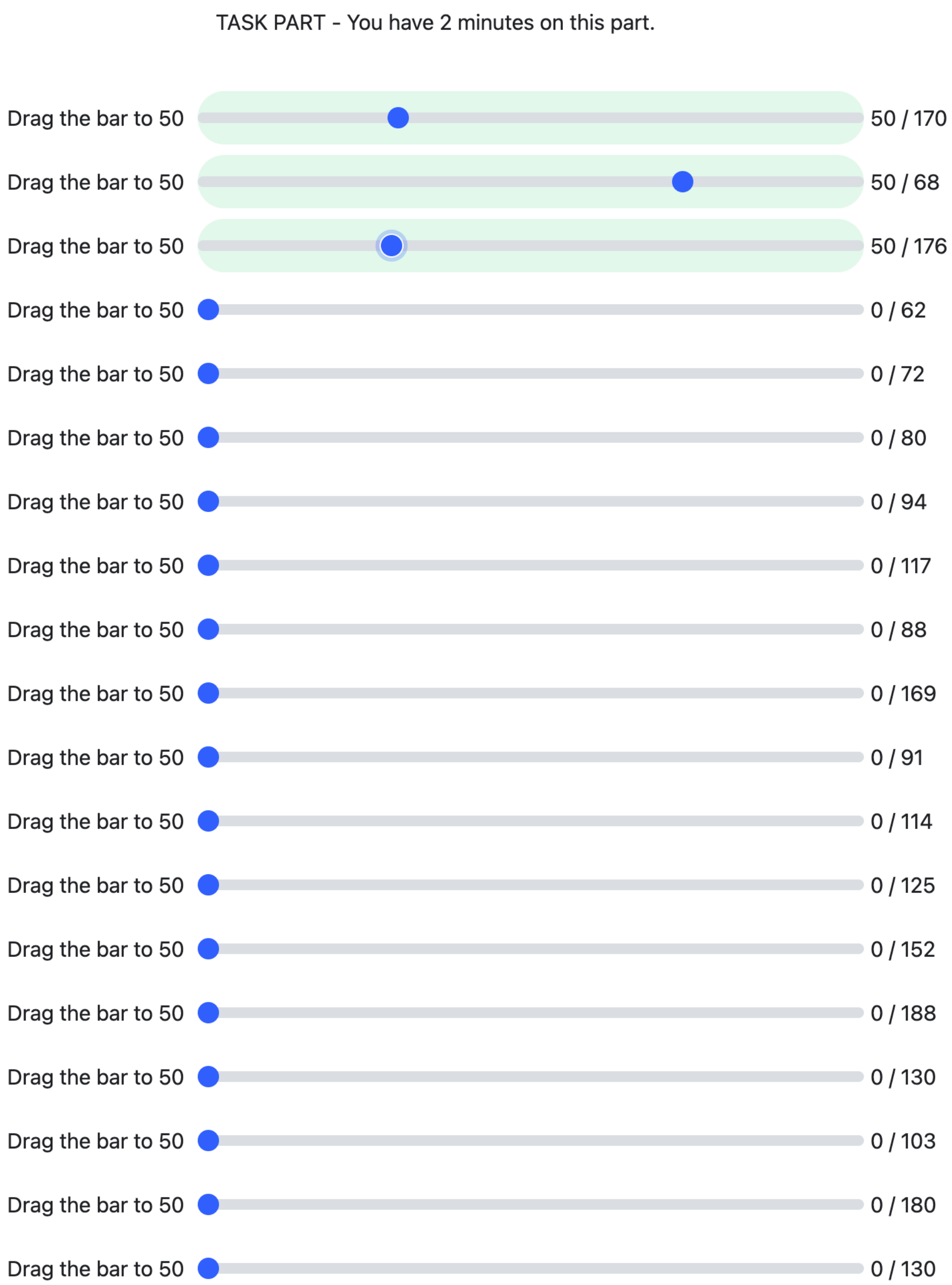


Figure A7: Slider Task

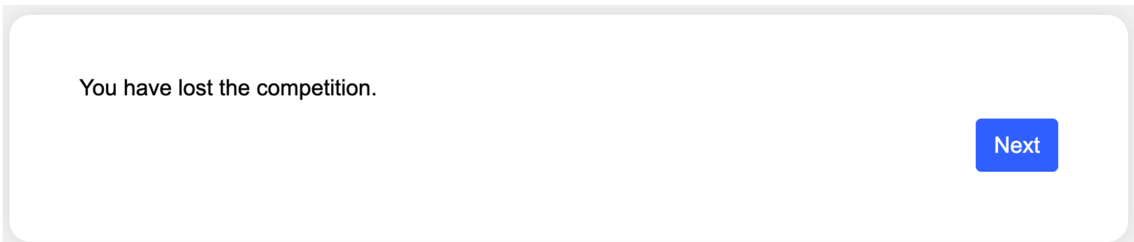


Figure A8: Feedback in the Slider Task

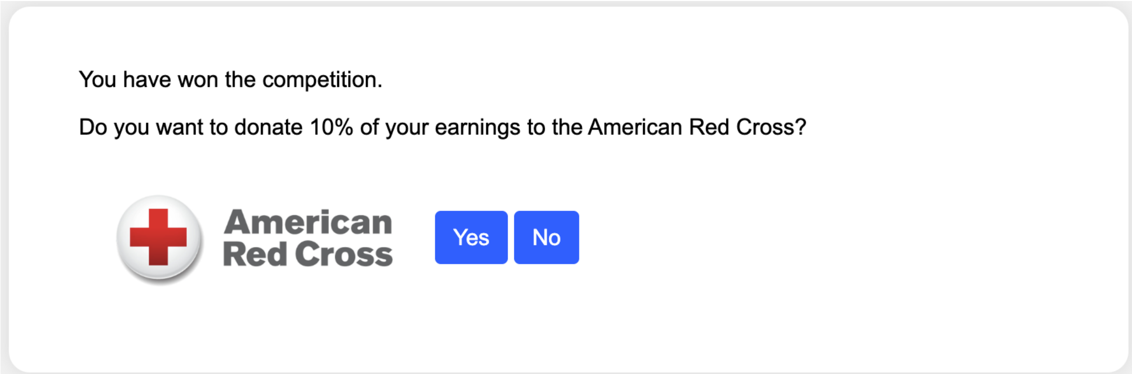


Figure A9: Feedback in the Slider Task: Winner Donation Decision

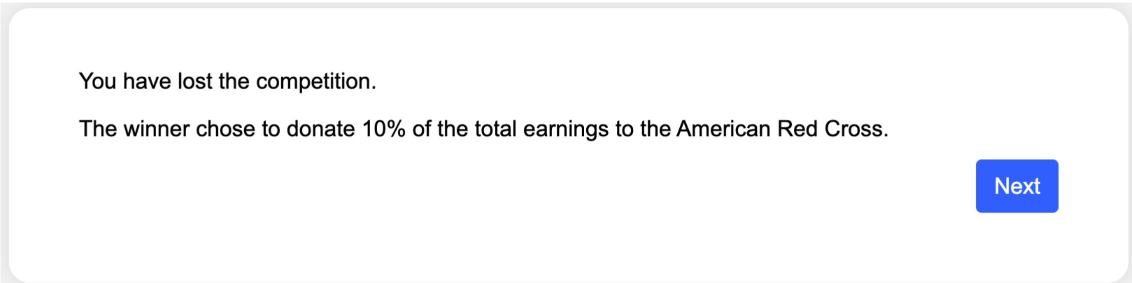


Figure A10: Feedback in the Slider Task: Loser Donation Signal

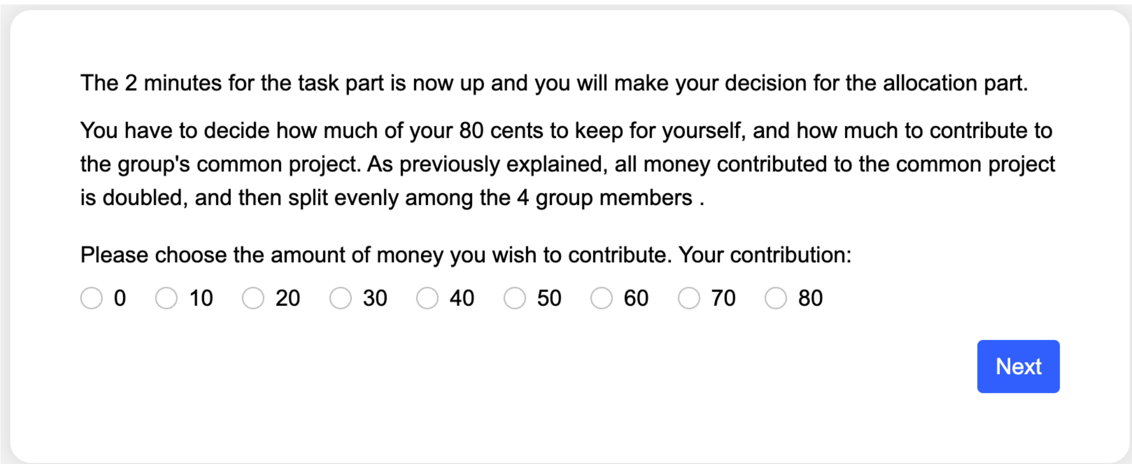


Figure A11: Cooperation Decision: Piece Rate

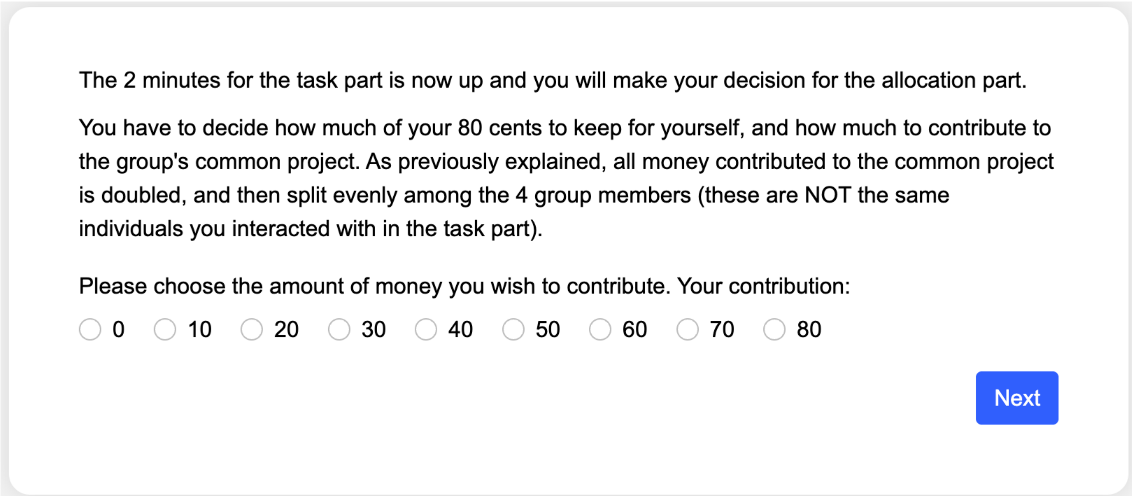


Figure A12: Cooperation Decision: Competition and Donation Treatments

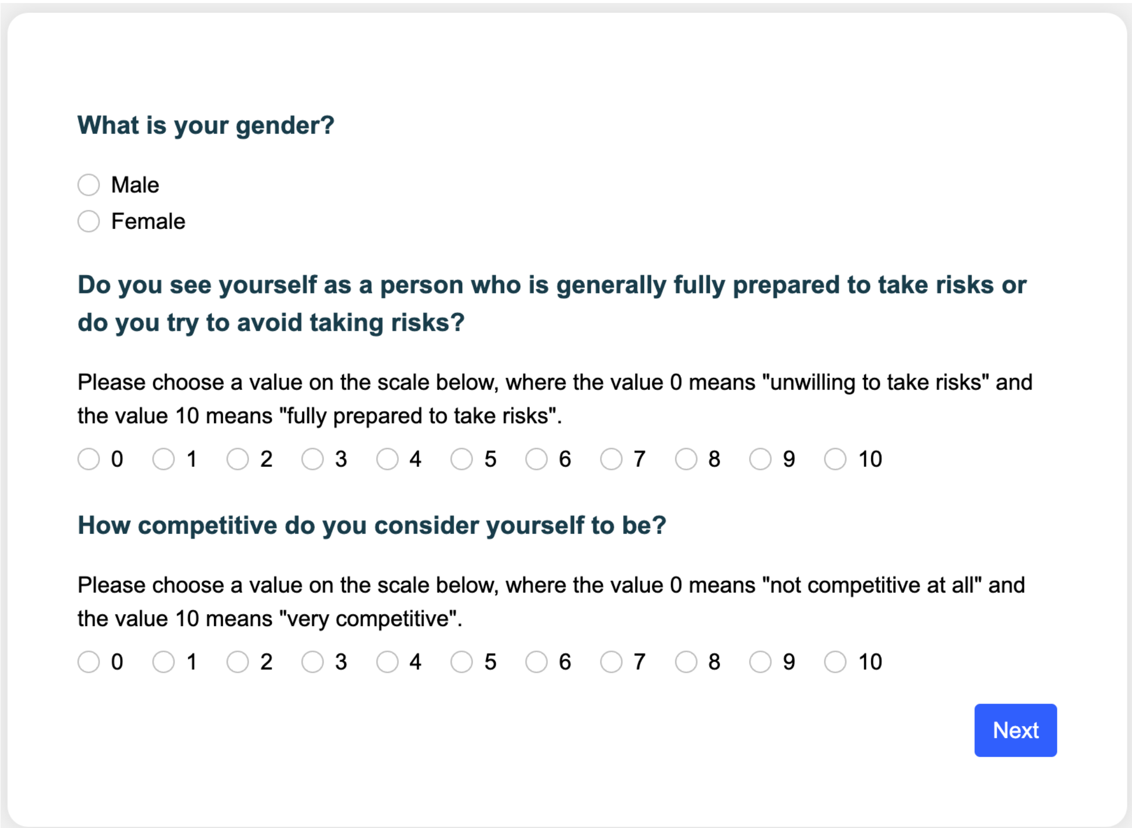


Figure A13: Final Survey

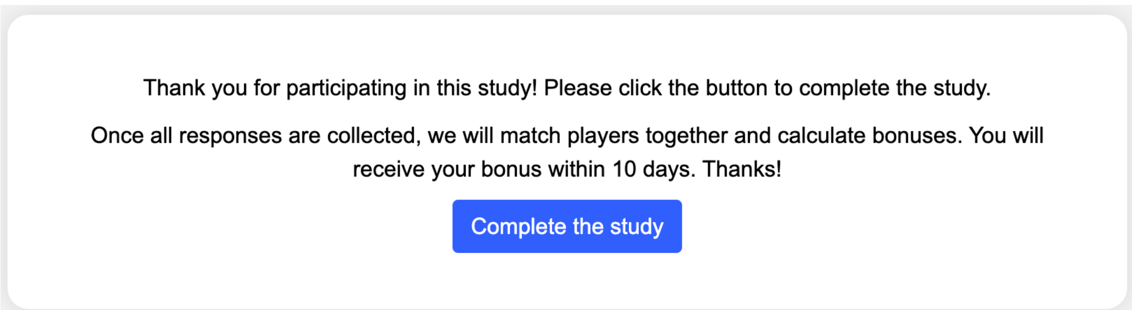


Figure A14: Study Completion

A.2 Instructions

Welcome

Thank you for participating in this study. You will receive $0.78 for participating if you pass the comprehension checks. You cannot participate in this study more than once. You also have the opportunity to receive additional money ($1.10 average bonus payment), which will be described in the next few pages.

General Instruction

You will participate in a "task part" and an "allocation part". You will be paid for 1 of the 2 parts - which one will be randomly decided after you have completed the study and is not influenced by your decisions. In the task part, you are paid according to your performance in a task. In the allocation part, you have to allocate money between yourself and 3 other randomly selected participants. All of you receive this same set of instructions. Please click the button to continue.

Slider Task Instructions: Piece Rate

In the task part, you will position sliders at the 50 mark, as in the example below: You will be given 2 minutes to correctly position as many sliders as you can. Your payment in this part depends on the number of sliders you position correctly. You will receive 4 cents per correctly positioned slider. You will not know the outcome of the task part today. Before we start, we will give you a chance to practice the task. There are 3 sliders below. Note that the maximum value differs between the sliders. You need to correctly position the 3 sliders on 50 in order to continue with the study. On the next page, you will receive more information about the allocation part. Please click the button to continue.

Slider Task Instructions: Competition

In the task part, you will position sliders at the 50 mark, as in the example below: You will be given 2 minutes to correctly position as many sliders as you can. Your payment in this part depends on the number of sliders you position correctly compared to 3 other participants. These are NOT the same individuals you interact with in the allocation part. If you correctly position more sliders than the other three participants, you will receive 16 cents per correct slider. If not, you will receive nothing. If there is a tie, one randomly selected participant among the top performers will be paid. You will know whether you won or lost immediately after the task part. Before we start, we will give you a chance to practice the task. There are 3 sliders below. Note that the maximum value differs between the sliders. You need to correctly position the 3 sliders on 50 in order to continue with the study. On the next page, you will receive more information about the allocation part. Please click the button to continue.

Slider Task Instructions: Competition + Donation

In the task part, you will position sliders at the 50 mark, as in the example below: You will be given 2 minutes to correctly position as many sliders as you can. Your payment in this part depends on the number of sliders you position correctly compared to 3 other participants. These are NOT the same individuals you interact with in the allocation part. If you correctly position more sliders than the other three participants, you will receive 16 cents per correct slider. If not, you will receive nothing. If there is a tie, one randomly selected participant among the top performers will be paid. You will know whether you won or lost immediately after the task part. If you win, you can choose to donate 10% of the prize to the American Red Cross. All group members will be informed about your donation decision. You will receive proof of the donation after the experiment. Before we start, we will give you a chance to practice the task. There are 3 sliders below. Note that the maximum value differs between the sliders. You need to correctly position the 3 sliders on 50 in order to continue with the study. On the next page, you will receive more information about the allocation part. Please click the button to continue.

Public Good Game Instructions

In the allocation part, each person in your group with 3 other people is given 80 cents (in addition to the 60 cents you received already for participating). You each decide how much of your 80 cents to keep for yourself, and how much (if any) to contribute to the group's common project (in increments of 10 cents: 10, 20, 30, etc). All money contributed to the common project is doubled, and then split evenly among the 4 group members. Thus, for every 2 cents contributed to the common project, each group member receives 1 cent. If everyone contributes all of their 80 cents, everyone's money will double: each of you will earn 160 cents. But if everyone else contributes their 80 cents, while you keep your 80 cents, you will earn 200 cents, while the others will earn only 120 cents. That is because for every 2 cents you contribute, you get only 1 cent back. Thus you personally lose money on contributing. The other people are REAL and will really make a decision - there is no deception in this study. Once you and the other people have chosen how much to contribute, the allocation part is over. Please, answer the following questions. Providing an incorrect answer twice will result in your exclusion from the study. After this, please click the button to start with the task part. Once the sliders appear the 2 minutes for the task part will start. What level of contribution earns the highest payoff for the group as a whole? What level of contribution earns the highest payoff for you personally?

Appendix B

B.1 Pre-Registration

Title: Competition and Cooperation: the Role of Donations (#177381)
Pre-registered on: 05/31/2024 01:00 AM (PT)
This pre-registration is currently anonymous to enable blind peer-review. It has one author.

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
We seek to replicate the effect of competition on cooperation in "Buser, T., & Dreber, A. (2016). The flipside of comparative payment schemes. Management Science, 62(9), 2626-2638." In particular, we aim to replicate two treatments of the original study, the piece rate and the feedback condition. Furthermore, we test the effect of two novel treatments. In these two conditions, we ask whether prosocial signals from winners can significantly reduce the negative spillover of competition on cooperation. We evaluate prosocial signals in the context of donation decisions, measuring the overall effect of donation schemes (embedding both prosocial and non-prosocial signals) on the generation of spillover effects from competition to cooperation. We further vary and test whether the strength of the signal plays a role.

3) Describe the key dependent variable(s) specifying how they will be measured.
The key dependent variables are the contribution in a standard four-player Public Good game (MPCR = 0.5) played after the competition and performance in a slider task (number of sliders solved).

4) How many and which conditions will participants be assigned to?
Each participant will only be assigned to one out of four experimental conditions. Two conditions are replications of the piece rate and feedback treatments in Buser, T., & Dreber, A. (2016). The additional two conditions are equal to the feedback treatment, except that winners can donate 10% (third condition) or 25% (fourth condition) of their prize to the American Red Cross. At the same time, losers observe the donation decision (on the feedback page).

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
The analysis will be based on the Stata Do-file uploaded by Buser, T., & Dreber, A. (2016) in the supplementary material of the paper we aim to replicate. Specifically, we will test differences in contributions to the public good game across all treatments using the Kruskal-Wallis test and Mann-Whitney test, distinguishing groups exposed to the prosocial and non-prosocial signals. We will use linear regression models to assess the effect of our treatment variations on contributions to the public good. Regression specification will include a dummy for each treatment detail condition (Feedback, Donation 10% (and 25%) with prosocial (and non-prosocial) signals). The piece rate condition will be excluded from the specification and used as a baseline comparison. We use the Wald test to assess differences in the coefficients, and more specifically, between the feedback condition, donation 10% (Prosocial) and Donation 25%

(Prosocial). Other control variables will be gender, performance, age, and risk attitude. We will report robust standard error in our specification. We will further test the treatment effect of the two donation programs by not distinguishing between prosocial and antisocial signals to comment on the policy implications of our treatment variations. Additionally, we will use a separate and pulled analysis of all prosocial signals (10% and 25%) and both donation programs (10%+25%) with respect to the Buser & Dreber (2016) feedback condition. Finally, a separate OLS model will test differences among participants exposed to the prosocial and non-prosocial signals within the subjects in both donation treatments. Gender differences will be reported, in line with the analysis conducted by Buser, T., & Dreber, A. (2016).

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

If sessions are interrupted due to technical issues, observations from those sessions will not be included in the analysis. Dropouts will be excluded (we test for selective attrition). Failure to correctly answer attention checks and control questions implies the exclusion from the study (they cannot proceed with the experiment). The study includes two attention checks and two control questions, equal across treatment conditions.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We chose the sample size using traditional a priori power analysis. We also ran simulations using the empirical distribution of contributions to the public good reported in Buser, T., & Dreber, A. (2016), which yielded the same sample sizes.

In order to be able to detect an effect size (Cohen's d) of 0,29 found in Buser, T., & Dreber, A. (2016), with 80% power and 5% statistical significance, we need a sample of 370 (185 in each group).

Regarding our two novel treatments, we will compare the contributions to the public good in each of these two treatments with the contributions in the feedback treatment. We expect that the possibility to donate reduces the negative effect of competition on cooperation only for the winners and losers who are exposed to the pro-social signal (i.e. to a donation by the winner). Without previous evidence on effect sizes we decided to set the minimal effect size (Cohen's d) we want to detect to 0,196 (65% reduction in the efficiency loss). To be able to detect such effect size with 80% power and significance level at 5% the total sample size needs to be 812 (406 in the feedback condition and 406 winners and losers exposed to the prosocial signal). As we expect that roughly half of the winners in the tournament will donate, we need to double the sample in the two novel conditions. Minimum detectable effect size with 80% power at a 5% statistical significance will be reported.

Therefore, we plan to recruit minimally a total of n = 2215 participants: 185 (piece rate), 406 (feedback), 812 (donation 10%), and 812 (donation 25%).

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We chose to conduct our study on Prolific rather than Amazon Mechanical Turk (MTurk), as in the original Buser, T., & Dreber, A. (2016). The reason stems from researchers now increasing sample requirements on MTurk to obtain higher quality data (>99% approval rate and >1000 HITs approved). While there's uncertainty about whether MTurk's quality has truly declined or is a matter of increased awareness of its limitations, the choice reflects a belief that Prolific represents a valid substitute to MTurk.

B.2 Session Timing

The 22 experimental sessions were conducted between 24 June 2024 and 24 July 2024. Sessions were distributed across seven different days, with most days hosting multiple sessions. Table B1 provides the detailed summary of each session, including session number, treatment, date, and participants.

Session #	Treatment	Date	Participants
1	Control	24 Jun 2024	9
2	Control	25 Jun 2024	31
3	Competition	25 Jun 2024	33
4	Control	02 Jul 2024	27
5	Competition	02 Jul 2024	93
6	Donation 10%	02 Jul 2024	167
7	Competition	03 Jul 2024	94
8	Donation 10%	03 Jul 2024	189
9	Donation 25%	03 Jul 2024	200
10	Control	04 Jul 2024	27
11	Control	04 Jul 2024	21
12	Control	04 Jul 2024	21
13	Competition	04 Jul 2024	91
14	Donation 25%	04 Jul 2024	185
15	Control	22 Jul 2024	18
16	Donation 10%	22 Jul 2024	239
17	Donation 25%	22 Jul 2024	207
18	Competition	23 Jul 2024	131
19	Donation 10%	23 Jul 2024	202
20	Donation 25%	23 Jul 2024	219
21	Control	24 Jul 2024	12
22	Control	24 Jul 2024	31

Table B1: Experimental Sessions

Each session was associated with one of the four treatments (Control, Competition, Donation 10%, Donation 25%). This scheduling ensured that treatments were evenly represented across the experimental period

Appendix C

C.1 Robustness

To account for potential interdependencies among participants sharing the same competitive experience in the first part of the experiment, we re-estimate the main specifications reported in Tables 2.2 and 2.5 clustering standard errors at the competition-group level. The results remain unchanged in magnitude and statistical significance.

	(1) Contribution	(2) Contribution	(3) Contribution
Competition	-0.974	-1.237	0.230
	(2.886)	(2.879)	(2.836)
Donation 10% Prosocial	1.063	1.051	1.817
	(3.179)	(3.189)	(3.169)
Donation 10% Non-Prosocial	-6.650**	-6.668**	-6.018**
	(2.804)	(2.807)	(2.768)
Donation 25% Prosocial	2.227	2.083	2.683
	(3.196)	(3.217)	(3.223)
Donation 25% Non-Prosocial	-5.863**	-5.782**	-4.967*
	(2.804)	(2.812)	(2.774)
male		-3.443**	-5.769***
		(1.508)	(1.530)
performance		0.336***	0.412***
		(0.086)	(0.086)
age		0.110**	0.117**
		(0.048)	(0.049)
risk			2.123***
			(0.288)
Constant	47.716***	37.109***	23.598***
	(2.400)	(3.816)	(4.170)
Observations	2247	2246	2246
R-squared	0.009	0.017	0.042

Clustered sandard errors at the competition group level in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C1: OLS Regression: Contributions in the Public Good Game (Cluster Robust)

	(1)	(2)	(3)	(4)	(5)	(6)
	Contribution	Contribution	Contribution	Contribution	Contribution	Contribution
Competition	2.339	-3.643	2.688	-4.173	4.205	-2.780
	(4.533)	(3.715)	(4.551)	(3.700)	(4.562)	(3.577)
Donation (Prosocial)	1.935	0.992	1.804	0.926	2.768	1.334
	(4.504)	(3.680)	(4.551)	(3.667)	(4.583)	(3.575)
Donation (Non-Prosocial)	-5.055	-7.132**	-4.900	-7.255**	-4.280	-6.418*
	(3.990)	(3.445)	(4.056)	(3.436)	(4.074)	(3.307)
performance			0.369***	0.298**	0.445***	0.365***
			(0.118)	(0.133)	(0.119)	(0.133)
age			0.083	0.122**	0.104	0.120**
			(0.101)	(0.052)	(0.100)	(0.054)
risk					1.800***	2.387***
					(0.448)	(0.388)
Constant	44.944***	50.000***	32.292***	38.788***	17.889**	24.477***
	(3.695)	(3.130)	(6.380)	(5.099)	(7.392)	(5.455)
Observations	987	1259	987	1259	987	1259
R^2	0.009	0.010	0.019	0.015	0.036	0.047

Clustered sandard errors at the competition group level in parentheses

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

Table C2: OLS - Gender (Cluster Robust)

Chapter 3

Misinformation and Emotions

A. Barbazeni

Abstract

This study investigates the relationship between misinformation and its emotional content, focusing on three dimensions: emotional intensity (overall emotional tone of the content), emotional valence (positive vs. negative emotions), and the emotional intensity associated to specific emotions (anger, fear, sadness, happiness, surprise, and disgust). Leveraging a novel classification method based on OpenAI's pretrained language models, we generate emotion scores for news content from both the perspective of an impartial media assistant and through the lens of Democrat and Republican voters. Our analysis reveals that emotional content is positively associated with misinformation, highlighting the importance of all emotions, but sadness. Furthermore, we observe heterogeneous patterns across simulated political perspectives, with Republican-leaning scores generally showing stronger emotional responses to misinformation than Democrat-leaning scores.

Keywords: Misinformation, Emotional Content, Emotional Valence, Polarization.
JEL Codes: D83, D72, C55, L82

3.1 Introduction

In the last decades, false and misleading information has rapidly spread, using different channels for their diffusion and manipulating various aspects related to its saliency (Gennaro and Ash, 2022; Grinberg et al., 2019; Vosoughi et al., 2018). Misinformation threatens the quality of public discourse and contributes to the formation of networks in which false news shapes beliefs (Djourelova et al., 2023; Lazer et al., 2018; Rathje et al., 2023), ultimately distorting public opinion (Kartal and Tyran, 2022). The need to detect and curb its spread is increasingly urgent, as ad-funded online news often exhibits lower quality (Liporace, 2021), with habitual sharers and echo chambers responsible for a large share of misinformation dissemination (Acemoglu et al., 2024; Ceylan et al., 2023).

A growing literature documents that false information spreads more widely than truthful content on digital platforms, largely driven by users' sharing decisions and their emotional reactions to information (McLoughlin et al., 2024; Vosoughi et al., 2018). Most existing studies therefore focus on the user side of information diffusion, examining how users emotionally respond to content and how these reactions shape sharing behavior. In contrast, much less attention has been devoted to the producers side of misinformation, namely to the emotional structure of the content itself. As a result, little is known about whether misinformation systematically differs from accurate information in terms of emotional intensity and composition, despite the possibility that information producers strategically embed emotional content to increase engagement and diffusion.

In this study, we investigate how the emotional content of information correlates with its accuracy and how this relationship varies with readers' political affiliation. Specifically, we investigate this relationship under three measures of the emotional content. First, we analyze the emotional intensity, defined as the overall degree of emotionality expressed in the content, regardless of the emotion's type. Highly emotional content, whether positive or negative, can lead to an increase in engagement, potentially affecting the spread of both accurate and false information. Second, we analyze the emotional valence, which refers to the emotional direction of information (positive or negative emotionality). In this regard, previous literature suggests that negatively valenced content spread faster, particularly in political contexts (Rasmussen et al., 2025; Soroka et al., 2019). Third, we examine the role of specific emotions, based on Ekman's emotion framework, focusing on emotions such as anger, fear, sadness, happiness, surprise, and disgust. We explore how each of these emotions are associated to accuracy. Through these three levels of analysis, we aim to uncover new patterns in how emotional framing affects accuracy judgments and dissemination of (mis)information, and how these effects vary depending on readers' political affiliation.

To do this, we leverage two complementary datasets: a comprehensive archive of fact-checked news headlines from PolitiFact.com (2010–2024), and a dataset of tweets from the Internet Research Agency (IRA) collected from 2014 to 2017. The first dataset contains a large number of headlines, each paired with an independently verified accuracy rating. However, reporters do not randomly choose which headlines to fact-check, resulting in a potential overrepresentation of false and emotional claims. The second dataset serves as a robustness check by providing an alternative source of misleading information that is not affected by reporters' selection. In contrast, the IRA dataset does not include an accuracy measure for tweets, so accuracy must be inferred from the trustworthiness of the source domain. We employ a novel emotional classification method, using the OpenAI platform's API to generate emotion scores for each headline. This results in an assignment mechanism that exploits a widely pre-trained language model capable of processing information about the underlying context. In contrast with other methods which often overlook the influence of broader context on emotional interpretation, our approach evaluates the headline as a whole, capturing its overall meaning and tone. Next, we analyze the correlation between the emotional content of headlines and their factual accuracy. We explore this relationship through three main levels of analysis: emotional intensity, emotional valence, and all the six basic Ekman's emotions (happiness, sadness, anger, fear, surprise, and disgust). Furthermore, we replicate our econometric model to assess how these emotional measures differ across political content and examine potential voters asymmetries in emotional perception using simulated scores. Specifically, we analyze whether the relationship between emotion and misinformation varies with the slant of the news or its audience.

We find that the emotional intensity and negative valence of information are positively correlated with inaccurate statements. Misinformation evokes emotions with higher intensity and it is framed more negatively than true news. These results are robust in both our datasets. We additionally provide suggestive evidence that this correlation varies depending on the reader's political affiliation. Our analysis indicates that negative emotions (anger, fear, and disgust) are strongly correlated with misinformation, while happiness is negatively associated with it. Interestingly, sadness does not emerge as a significant predictor, suggesting it plays a different role in how information is perceived. Furthermore, surprise positively correlates with inaccuracy, indicating that information that evokes unexpected content may increase susceptibility to falsehoods. Finally, we show that our measures of emotional content increase over the years, and that social media and other internet sources exhibit higher levels of emotional intensity and valence than traditional sources.

Our study makes several contributions to the existing literature. First, we provide robust empirical evidence of the relationship between emotional content and misinformation, covering a broader spectrum of emotions and focusing on the emotionality featuring the content, rather than users' reactions to misinformation. Second, we introduce an innovative emotion classification method that enhances the semantic understanding of emotionality in text. Third, we do not limit our analysis to information stemming from social media, expanding the set of sources to include traditional media (e.g. TV, Events, Press, etc.). Finally, we explore ideological asymmetries by analyzing heterogeneous correlations across voters, shedding light on how emotional content reinforce in-group bias and out-group animosity in political contexts.

Emotions play a central role in how misinformation spreads and is consumed online. The existing evidence suggests that false news tends to spread more widely and quickly than true news on social media, often because it elicits stronger emotional reactions such as fear, disgust, and surprise in user replies (Vosoughi et al., 2018). This dynamic is further reinforced by a general negativity bias, which has been documented across various countries (Soroka et al., 2019), increasing engagement, especially when it targets out-groups (Rasmussen et al., 2025). Emotional language, especially outrage, is a key driver of misinformation dissemination, and users frequently share outraged content without even reading it (McLoughlin et al., 2024). Moreover, emotions have been found to heighten belief in false, but not true, news stories (Martel et al., 2020).

Furthermore, social media platforms amplify out-group hostility, as posts targeting political opponents are shared more often and elicit greater engagement through anger, while in-group language causes a larger amount of "love" reaction (Rathje et al., 2021). Political discourse has also been experiencing an increase in emotionality since the late 1970s, particularly among Democrats and ideologically extreme members, reflecting a broader trend toward affective polarization (Gennaro and Ash, 2022). In addition, motivated reasoning plays a crucial role in how individuals process information, contributing to belief biases, polarization, and overconfidence in one's views (Thaler, 2024). These dynamics are exacerbated by heterogeneous reactions to events and information. For example, natural disasters can polarize climate change beliefs, amplifying concern among liberals while reducing it among conservatives (Djourelova et al., 2023). Similarly, misinformation has varying effects across populations: older adults, conservatives, and politically engaged individuals are disproportionately responsible for fake news sharing, especially during election cycles (Grinberg et al., 2019); this pattern is not merely due to ignorance, but often reflects partisan animosity and a desire to discredit political opponents (Osmundsen et al., 2021). Despite these risks, analytical thinking and Democratic affiliation have been associated with improved ability to discern true from false news, while ideological alignment and familiarity tend to bias truth judgments (Sultan et al., 2024).

Misinformation remains a significant societal concern due to its effects on political behavior, public health, and trust in institutions (Barrera et al., 2020; Nyhan and Reifler, 2015; Pennycook et al., 2020). Although people show some ability to distinguish true from false information (Pennycook and Rand, 2019; Pfänder and Altay, 2025), they often struggle to detect deception and remain overconfident in their judgments (Serra-Garcia and Gneezy, 2021). On one hand, interventions such as accuracy prompts, financial incentives, and fact-checking can improve judgment and reduce sharing of false news (Burdea and Woon, 2023; Henry et al., 2022; Pennycook et al., 2021; Rathje et al., 2023). On the other hand, we necessitate more producer-targeted interventions to better understand the supply-side drivers of false news generation. In this regard, Serra-Garcia (2025) offer an important insight: producers' incentives to attract the attention of information receivers lead them to transmit less complete information, but not necessarily more inaccurate information. According to Serra-Garcia (2025), distorted information and beliefs varies across individuals depending on their willingness to engage more deeply with a topic: more engaged users tend to acquire more accurate knowledge, whereas less engaged users are more likely to hold distorted beliefs. This finding is crucial, as it provides a direction for investigating the mechanisms underlying the spread of misinformation and helps explain why misinformation tends to be more emotionally engaging than accurate information.

Our findings suggest that policies and interventions improving discernment should account for the emotional content of misinformation. Targeting emotionally charged content may help reduce its spread, by designing platforms to flag emotional content and by promoting emotional literacy. Moreover, conditional on the identification of non-trustworthy sources, platforms might reconsider amplification algorithms that favor emotionally provocative content.

The remainder of the paper is structured as follows: Section 2 provides an overview of the methodology used and data, Section 3 reports the results, and Section 4 concludes.

3.2 Data and Methodology

Our main source of data originates from the PolitiFact website (https://www.politifact.com), an independent fact-checking platform based in Florida. PolitiFact functions as a neutral monitor of political discourse, evaluating the accuracy of publicly stated claims made by politicians, activists, talk show hosts, and economists across various media (e.g., speeches, newspapers, television). Each day, PolitiFact identifies and evaluates statements that can be objectively verified, claims that assert facts rather than express promises or opinions, which cannot be checked using empirical data.

We collect headlines from the beginning of PolitiFact's archives, from 2007 to 2024 (N=23474). However, due to the limited numbers of observations, we exclude the years 2007-2009[1]. PolitiFact rates the accuracy of each claim using the "Truth-O-Meter", a six-level scale measure of accuracy: True (the statement is accurate and nothing significant is missing), Mostly True (the statement is accurate but needs clarification or additional context), Half True (the statement is partially accurate but leaves out important details or takes things out of context), Mostly False (the statement contains an element of truth but ignores critical facts that would give a different impression), False (the statement is not accurate), and Pants on Fire (the statement is not accurate and makes a ridiculous claim). Furthermore, each rating is accompanied by a rationale and sources used in the assessment. We use web scraping to collect data of the statement analyzed by Politifact, its date, author, source, and the corresponding reporter. These variables will be included in the regression specification as controls to account for potential confounding factors. Finally, we construct a Misinformation Index by reversing the "Truth-O-Meter" scores.

We use Politifact's statements to classify our data, generating emotional scores on a seven-point Likert scale for each of the six Ekman's basic emotions. Using platform OpenAI, we built a prompt requiring the generation of a vector of values associated with each of the six emotions. The role we assign to the model within the system prompt states "You are an expert media assistant with a psychological background.", which aims at providing the most neutral evaluation of each statement. To assure that values were not generated randomly, we run a small pilot exercise in which we additionally ask to provide a reasoning behind the value assignment and assess its consistency (details in Appendix). Other three prompt sets were used to classify the statements. The first two consist of redefined version of the original one, with a difference in the system prompt used. We replaced the prompt "You are an expert media assistant with a psychological background." with "You are a US democratic voter." and "You are a US republican voter.", obtaining two other proxies for emotional intensity. Finally, a third prompt

[1]The years 2007–2009 account for 1133 observations. Including these observations in the analysis does not change the results presented in the following section.

was used to classify the political slant of the news, allowing us to identify pro-republican and pro-democrats news. We calculated our proxies for measuring the emotional content of the headlines in our dataset as follows. To compute emotional intensity, we use both weighted and simple average calculations across all emotion types[2]. The first measure (Emotionality (weighted)) sums positive and negative emotions and surprise (which is neither positive nor negative), weighting each group by one-third:

$$\text{Emotionality (weighted)} = \frac{1}{3}\times\text{happiness}+\frac{1}{3}\times\text{surprise}+\frac{1}{3}\times\frac{1}{4}\times(\text{sadness}+\text{anger}+\text{fear}+\text{disgust})$$

To enhance the robustness of our measure of emotional intensity, we also consider a model containing the simple average (Emotionality (simple)) across all emotions:

$$\text{Emotionality (simple)} = \frac{1}{6}\times(\text{happiness}+\text{sadness}+\text{anger}+\text{fear}+\text{surprise}+\text{disgust})$$

Regarding the index used to analyze the direction (positive or negative) of the effect of emotionally intense headlines, we construct the emotional valence variable by subtracting the average of all negative emotions from happiness (Fiala and Noussair, 2017). "Surprise" is therefore excluded from this computation:

$$\text{Positive Emotional Valence} = \text{happiness} - \frac{1}{4}\times(\text{sadness}+\text{anger}+\text{fear}+\text{disgust})$$

Among the controls used in our models, we include years dummies to control for time-specific variation in our measures of the emotional content. Figure 3.1 shows how the emotional intensity of headlines has increased during the last decade, accompanied by a decrease (increase) in the positivity (negativity) featuring the content.

In Figure 3.2, we report the coefficients obtained from Spearman's correlation test on our main variables. All variables measuring emotional content show a strong association with the Misinformation Index. Moreover, the moderate correlation between Emotionality (weighted) and Emotional Valence indicates that these variables capture related but distinct dimensions, consistent with our goal of generating separate measures of intensity and direction of the emotional content (positive or negative).

A second dataset was used to perform a robustness check on possible selection issue with Politifact's statements. Because Politifact's reporters endogenously choose which news to evaluate, the analysis may suffer from selection bias.

According to PolitiFact's editorial methodology, journalists do not evaluate a random sample of statements but instead select claims based on several criteria, including whether the statement is verifiable, widely circulated, or appears potentially misleading. As a result, the probability that a claim is fact-checked may correlate with characteristics such as its political salience, its likelihood of being false, or its potential to attract attention. Moreover, despite PolitiFact's

[2]The weighted index assigns equal weight to positive emotions (happiness), neutral emotions (surprise), and negative emotions (sadness, anger, fear, disgust), thereby eliminating valence-related asymmetries in the final score. Surprise is treated as neutral, as it is not inherently positive or negative (Baudouin et al., 2025). The simple average across all six basic emotions provides a complementary, assumption-free measure, serving as a robustness check against the weighting scheme, and can better account for negativity bias.

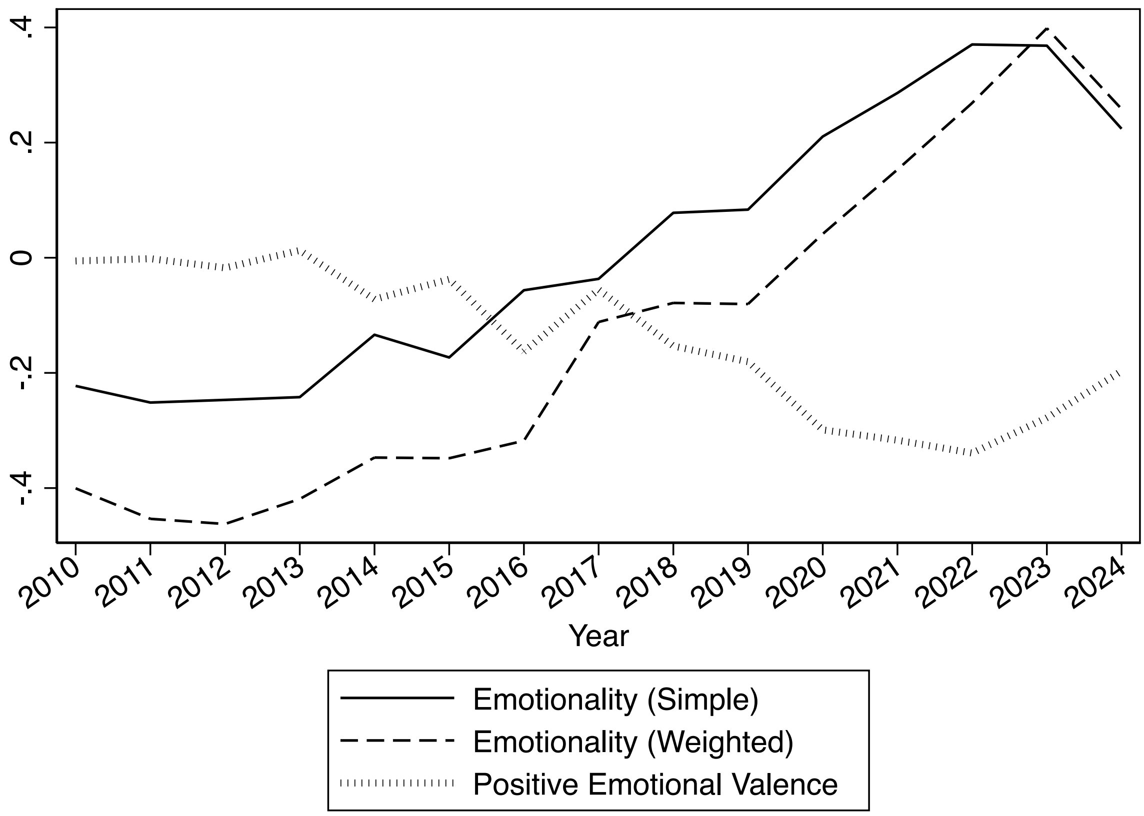


Figure 3.1: Emotional content over time

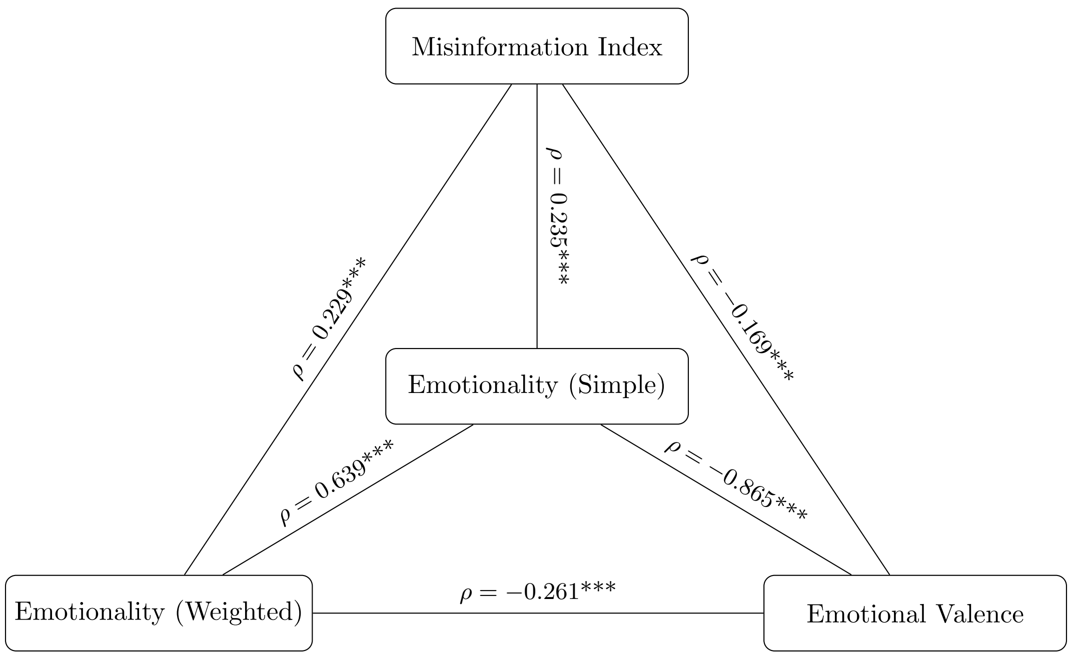


Figure 3.2: Spearman correlation between the Misinformation Index and Emotional Measures

commitment to selecting a balanced sample of claims from Republicans and Democrats, the salience of political topics may also lead to an overrepresentation of emotionally charged claims.

We therefore acknowledge that our estimates reflect correlations that may be biased by the overrepresentation of false, political, and emotional claims. To address potential concerns that our results are driven by this selection process, we complement the analysis with a second dataset based on tweets from the Internet Research Agency (IRA), which provides an alternative source of misleading content that is not generated through the fact-checking selection mechanism[3]

We collected data of the Internet Research Agency (IRA), a comprehensive dataset of tweets highlighting the presence of misinformation among American politicians. We focus our attention on tweets with URL links published from 2014 to 2017[4]. Following McLoughlin et al. (2024), we expanded the tweets' link, extracted their domain source (e.g., www.nytimes.com) and match them with the domain dataset provided by the authors. The domain dataset consists of a list of domain matched with a dummy variable classifying the source as "Trustworthy" or "Misinformation". Furthermore, using the Python library "BeautifulSoup", we extracted the headlines of the tweets' link and repeated the analysis of classification for each emotion under the role of a expert media assistant. Crucially, the unit of analysis in the IRA robustness check is the article content linked in each tweet, not the tweet text itself. We successfully extracted and classified 18718 observations. Differently from McLoughlin et al. (2024), we do not download tweet responses, but we focus on scraping the content linked in the tweets URL and perform text analysis directly on that content.

Several limitations of the IRA robustness check deserve explicit acknowledgement. First, the IRA dataset does not provide article-level accuracy ratings: misinformation status is inferred from domain-level source quality, which is a less precise measure than Politifact's statement-level Truth-O-Meter. Second, a substantial share of the linked URLs contained deleted or otherwise inaccessible content at the time of scraping, so the extracted sample is not fully representative of all content shared by IRA accounts. For these reasons, the IRA results should be interpreted as supporting the main findings rather than as an independent replication. Taken together, the two datasets complement one another: Politifact offers precise, statement-level accuracy ratings across a wide range of media sources, while the IRA dataset provides a plausibly exogenous source of misleading content that is free from the fact-checking selection mechanism (results reported in Appendix).

To compare the two datasets, we generated a dummy variable for Politifact's Truth-O-Meter, classifying as "Misinformation" news labeled as "Mostly False", "False" or "Pants on Fire", while deeming "Trustworthy" all news indicated as "Mostly True" and "True".

Table 3.1 presents the descriptive statistics for the key variables used in our empirical analysis. The last three columns report the means and difference of scores assigned by OpenAI when prompted to act as a Republican or Democratic voter. Overall, the average emotionality score is slightly higher when measured through partisan lenses, supporting the evidence that partisans perspectives may affect judgment. With the exception of "Disgust", all Republican vs Democrat differences (obtained from simulated emotional scores) are statistically significant at the

[3]In general, empirical analyses of misinformation typically rely on non-representative samples of news. Some datasets are constructed from specific platforms or contexts (e.g., social media archives), while others lack a measure of factual accuracy. For these reasons, we employ two complementary datasets to assess whether the relationship between emotional content and truthfulness is consistent across different data sources and studies, rather than aiming to replicate the exact same empirical estimates.

[4]Data outside this range are no longer public.

1% level (t-test), leaving room for exploring differences across opposite political preferences[5]. Notably, valence scores are on average negative, suggesting a prevalence of negatively charged statements in the dataset. The average Misinformation Index is 3.88, indicating a tendency toward partially true or misleading claims. Furthermore, around 72% of statements (excluding half-true statements) are classified as factually false.

Variable	Mean	SD	Min	Max	Mean (Rep, R)	Mean (Dem, D)	Mean (R-D)
Happiness	1.88	1.35	1.00	7.00	2.43	2.13	0.30***
Sadness	3.09	1.33	1.00	7.00	3.30	3.73	-0.43***
Anger	3.70	1.45	1.00	7.00	4.25	4.22	0.03***
Fear	3.27	1.36	1.00	7.00	3.48	3.74	-0.25***
Surprise	3.53	1.08	1.00	7.00	3.75	3.81	-0.06***
Disgust	3.45	1.45	1.00	7.00	4.07	4.07	-0.00
Emotionality (weighted)	2.93	0.54	1.00	5.42	3.32	3.30	0.02***
Emotionality (simple)	3.15	0.76	1.00	5.83	3.55	3.62	-0.07***
Emotional Valence	-1.50	2.39	-5.75	6.00	-1.34	-1.81	0.47***
Misinformation Index	3.88	1.56	1.00	6.00			
False	0.72	0.45	0.00	1.00			
Observations	23474						

Result from t-test: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.1: Summary statistics

Finally, the pie chart in Figure 3.3 illustrates the distribution of news sources by percentage. Social media is the most prominent source, accounting for 33.98% of the total, followed by TV/Radio at 23.21% and events at 21.20%. Other internet sources contribute 11.91%, while traditional press makes up the smallest share at 9.70%. The source analysis reveals substantial differences in emotional content across sources. Figure 3.4 reports the mean values of our three measures of emotional content for each source type. For graphical purposes, we plot the negative emotional valence, which represents the negativity of the statement. Across media types, we observe that traditional sources such as TV, press, and events exhibit lower levels of both emotional intensity and negativity. In the next section, we include controls for source effects in our regression specification.

[5] Because scores can capture biases within the training sources of OpenAI, our aim is to extract insights that can inform future experimental designs on the causal effect of emotional content on misinformation, and to guide the experimental literature on the development of testable hypotheses regarding heterogeneous effects across partisan groups.

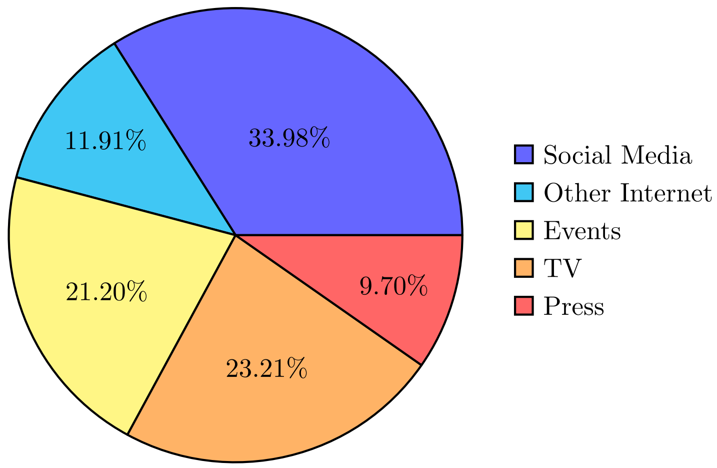


Figure 3.3: Distribution of sources by percentage

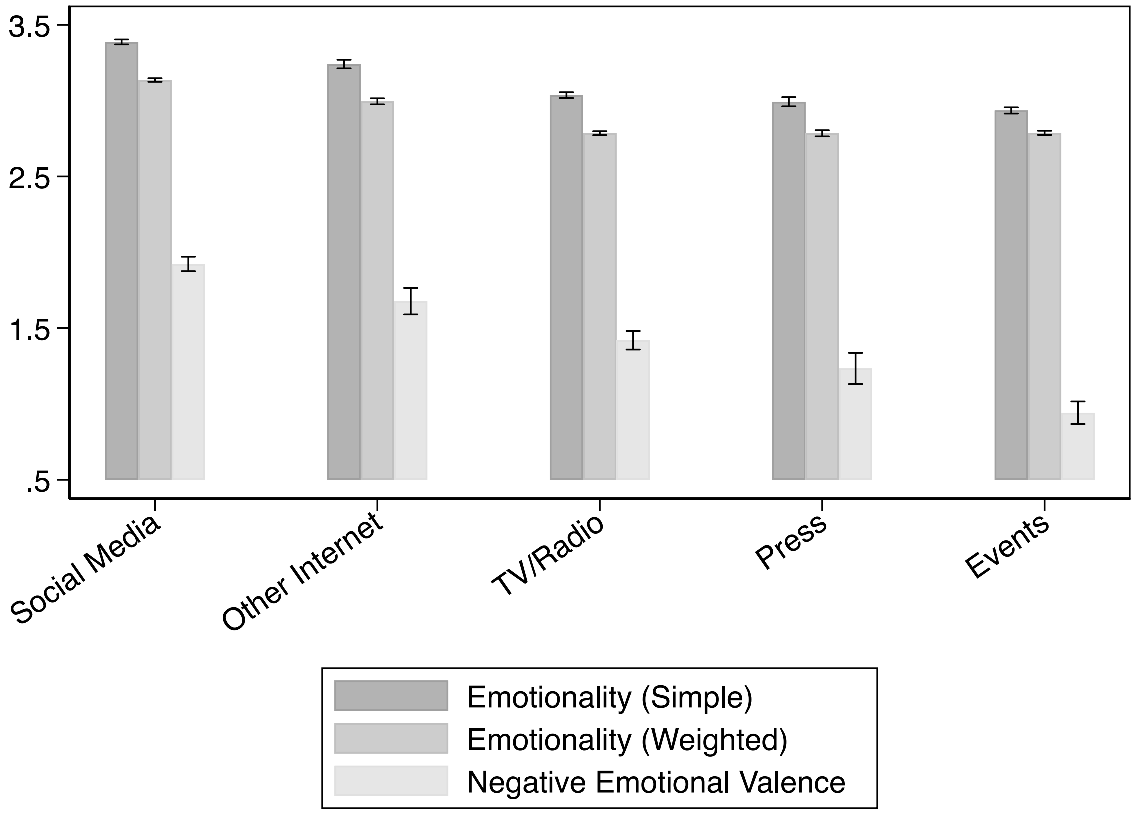


Figure 3.4: Emotionality and Sources

3.3 Results

In this section, we explore the association between misinformation and emotional content across all fact-checked statements in our dataset. Using data from the Politifact dataset, we estimate several regressions models where the main independent variable is the Truth-O-Meter score assigned by reporters publishing on the Politifact platform. Specifically, we consider the inverse of the Truth-O-Meter (Misinformation Index) as measure of the falsehood level of a given statement. Using sets of binary variables, we include fixed effects for year, season, frequent author of the statement (>100 statements made), and frequent reporters (>500 statements analyzed). To ensure comparability, all variables measuring emotional content and the Misinformation Index are standardized to have mean zero and standard deviation equal to one. All effects can therefore be interpreted as standard deviation units.

3.3.1 Emotional Intensity and Valence

We start our analysis by observing how emotional content is correlated with misinformation in two respects. First, we consider the overall emotional intensity of information, and second, determine whether there exists an asymmetric effect among positive and negative emotional content by studying the correlation between the falsehood of information and its emotional valence. Based on results from previous studies, we expect emotional intensity to be a valid predictor of inaccuracy (Martel et al., 2020; Vosoughi et al., 2018), and due to the presence of negativity bias, we expect misinformation to be negatively valenced (Rasmussen et al., 2025; Soroka et al., 2019). We first analyze the emotional scores computed by OpenAI APIs when prompted with instructions "You are an expert media assistant with a psychological background". Table 3.2 reports the estimates of three models with different dependent variables: weighted average emotionality (1), simple average emotionality (2), and emotional valence (3)[6]. Across all models, we observe that more inaccurate statements are significantly correlated with its emotionality. In particular, we observe that as the level of inaccuracy increases, emotional intensity rises (column (1): $\beta = 0.099$, $p < 0.01$; column (2): $\beta = 0.133$, $p < 0.01$). Additionally, column (3) shows that the association between emotionality and accuracy is primarily explained by negative emotional content: the more inaccurate a statement is, the less positive its emotional valence (column (3): $\beta = -0.120$, $p < 0.01$). This indicates that misinformation is not only more emotional but also more negatively framed, suggesting that misinformation producers might intentionally exploit negativity biases to increase the spread of false contents. Additionally, when excluding the Misinformation Index, the R-squared of the models decreases slightly, dropping from 0.124 (1), 0.106 (2), and 0.071 (3) to 0.117, 0.091, and 0.056 respectively (see Table A5 in Appendix A.2), suggesting that the Misinformation Index contributes modestly but consistently to the models' explanatory capacity. In table 3.2, we further report the coefficients for the Misinformation Index when emotional intensity and valence are calculated using the scores assigned by using the reader's perspective of Republicans and Democrats (Extra Models, see Appendix A.1 for fully specified models). The results anticipate that partisans' differences can provide useful insights on potential heterogeneous reactions to misinformation, which we explore later in the paper.

Media source effects are also notable: statements originating from traditional media sources such as TV, radio, press, or live events tend to show lower levels of emotionality and slightly more positive valence compared to social media sources (omitted category). This suggests that misinformation from traditional media is more trustworthy and might be subject to less

[6]The weighted average emotionality is calculated as (1/3) * happiness + (1/3) * surprise + (1/12) * (sadness + anger + fear + disgust). The simple average emotionality is the unweighted mean of happiness, sadness, anger, fear, surprise, and disgust. The valence of emotions is calculated as happiness minus (1/4) times the sum of sadness, anger, fear, and disgust.

manipulative intents.

	(1) Emotionality (weighted)	(2) Emotionality (simple)	(3) Emotional Valence
Misinformation Index	0.099***	0.133***	-0.120***
	(0.007)	(0.007)	(0.006)
`Extra Models`			
Republicans†	0.143***	0.154***	-0.118***
Democrats†	0.059***	0.110***	-0.137***
Other Internet	-0.007	0.014	-0.008
	(0.024)	(0.023)	(0.022)
Events	-0.310***	-0.252***	0.149***
	(0.022)	(0.021)	(0.020)
TV/Radio	-0.342***	-0.187***	0.039**
	(0.021)	(0.020)	(0.019)
Press	-0.279***	-0.162***	0.055**
	(0.026)	(0.026)	(0.024)
Constant	-0.130***	-0.051	-0.088***
	(0.033)	(0.032)	(0.030)
R^2	0.124	0.106	0.071
Observations	23174	23174	23174

Standard errors in parentheses
Fixed effects included: Year, Season, Author, Reporter
†Coefficients from model with emotions scores computed from a partisan perspective
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.2: OLS Emotionality & Valence

Next, specific emotions may correlate with different magnitude with inaccurate content, and understanding which best predict false information can help inform more effective policy interventions. For instance, if anger drives out-group animosity, targeted strategies could prompt reflection on the polarizing content. Conversely, if the surprising nature of misinformation increases user engagement on social media, policies that moderate or regulate algorithmic amplification based on engagement could be implemented to reduce its spread. To explore the correlation between misinformation and single emotions and account for possible correlation of the error term across different models, we employ seemingly unrelated regression (SUREG) model. Table 3.3 shows that inaccurate statements are strongly associated with heightened emotional responses across all range of emotions, but sadness. Specifically, the more inaccurate a statement is, the less likely it expresses happiness and the more likely correlates with increasing levels of anger, fear, surprise, and disgust. All coefficients are significant and report an effect size ranging from 0.107 to 0.158 standard deviations. Among them, fear, surprise and disgust show the strongest associations with misinformation in the data, while sadness may not represent one of the channels that affect the generation of false news. As in the previous table, we report the coefficients related to the models using partisans score of emotions to detect possible differences that can be informative for future research (full models in the Appendix). In the next subsection, we address partisans' differences in detail.

Control variables such as media source types yield expected patterns; for example, traditional media sources like press and TV/radio are generally associated with lower expressions of emotion

	Happiness	Sadness	Anger	Fear	Surprise	Disgust
Misinformation Index	-0.107***	-0.000	0.128***	0.151***	0.158***	0.145***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
`Extra Models`						
Republicans†	-0.103***	0.013*	0.133***	0.148***	0.180***	0.154***
Democrats†	-0.153***	0.024***	0.105***	0.126***	0.167***	0.130***
Other Internet	0.003	-0.010	0.037*	0.021	-0.039*	0.019
	(0.021)	(0.022)	(0.023)	(0.023)	(0.023)	(0.023)
Events	0.109***	-0.081***	-0.185***	-0.140***	-0.415***	-0.215***
	(0.020)	(0.021)	(0.021)	(0.021)	(0.022)	(0.021)
TV/Radio	-0.004	-0.095***	-0.051**	-0.095***	-0.411***	-0.064***
	(0.019)	(0.020)	(0.020)	(0.020)	(0.021)	(0.020)
Press	0.026	-0.007	-0.083***	-0.086***	-0.363***	-0.112***
	(0.024)	(0.025)	(0.025)	(0.026)	(0.026)	(0.025)
Constant	-0.161***	-0.140***	0.098***	-0.103***	0.050	0.101***
	(0.029)	(0.031)	(0.031)	(0.032)	(0.032)	(0.032)
R^2	0.055	0.024	0.072	0.096	0.194	0.083
Observations	23174					

Standard errors in parentheses
Fixed effects included: year, quarter, frequent authors, frequent reporters
†Coefficients from model with emotions scores computed from a partisan perspective
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.3: SUREG Specific Emotions

compared to social media sources.

To assess whether our findings are driven by the editorial selection process of Politifact's, we conduct a robustness analysis using a supplementary dataset based on tweets from the Internet Research Agency (IRA). This dataset provides an alternative source of potentially misleading content that is not generated through the fact-checking selection mechanism. Overall, the results obtained from the IRA dataset are consistent with those from the Politifact analysis (Appendix A.2). This suggests that the positive association between misinformation and emotional content is not solely driven by the criteria used by fact-checking organizations to select claims for verification. Rather than eliminating all potential sources of selection, this robustness exercise provides additional evidence that the relationship between emotionality and misinformation emerges across different data sources and empirical settings, in line with previous findings analyzing users' emotional responses to misinformation (McLoughlin et al., 2024; Vosoughi et al., 2018).

Overall, these findings provide a better understanding in assessing the correlation between emotional content and misinformation, addressing the role of specific emotions when associated with false content. In the next sub-section, we explore whether such effects are also relevant when assessing partisans' differences in emotional response.

3.3.2 Simulating Republican and Democrat Reactions to Misinformation

The next step for our analysis concerns the detection of differences in partisans' simulated emotional scores, which allow us to explore potential heterogeneous correlations, providing informative results about the drivers of misinformation among partisans, and leading the formulation of hypotheses for future studies. Assessing the relationship between emotions and the accuracy of statements using OpenAI APIs offers the advantage of generating scores based on the overall meaning and context of a statement. By adjusting the prompts used, we can also gain insights into how the model assigns these scores. In our dataset, we observed that the prompt "You are an expert media assistant with a psychological background" tended to assign moderate emotional scores to some of the statements involving political topics, irrespective from the emotion being positive or negative. We hypothesized that this prompt may have evaluated political content from a balanced perspective, considering viewpoints from both Republicans and Democrats. To address this, we reclassified our data using two additional prompts: "You are a US Republican voter" and "You are a US Democratic voter". The latter allow us to assess whether the emotional impact of misinformation varies across political preferences. Importantly, these scores do not reflect actual psychological reactions, but represent proxies stemming from the training process of OpenAI. Differently from the previous analysis, partisans' scores can be more subject to biases and their use should be limited to the formulation of hypotheses for future studies.

In tables 3.2 and 3.3, we already provided a preview of the coefficients for the Misinformation Index when the emotional scores were computed using Republicans and Democrats' perspectives, suggesting that partisans' emotional response to misinformation might differ across political affiliations (see the Appendix for further details). Here, we replace the outcome variable with the differences in emotional responses among partisans. This delta (Republican-Democrat; henceforth "R-D") represents the difference in emotional scores assigned to Republicans and Democrats for the same content. Positive values indicate stronger (weaker) emotional responses among Republicans (Democrats) compared to Democrats (Republicans). The results are reported in Table 3.4 (raw difference), Table 3.5 (absolute value of the difference) and Table 3.6 (raw difference of specific emotions), each table capturing different proxies for emotional content.

Table 3.4 presents estimates for the difference between Republicans and Democrats in overall emotionality (weighted), emotionality (simple), and emotional valence. The positive and statistically significant coefficient on the Misinformation Index in columns (1) and (2) suggests that Republicans experience greater emotional intensity compared to Democrats when exposed inaccurate statements. The opposite holds for Democrats, as shown by the coefficients in Table 3.2, reporting a lower magnitude compared to those computed using the standard prompt. This result is consistent with the differences reported in the previous tables: misinformation is associated with stronger (weaker) emotional reactions among Republicans (Democrats). The significant positive effect in emotional valence differences (column 3) also shows that inaccurate information lead to either an increase in positive valence for Republicans, an increase in negative valence among Democrats, or both. However, the coefficient is small and may indicate that the direction of the effect of misinformation on emotions is negligible when observing between-partisans differences.

Overall, these findings point to an asymmetry in simulated responses to misinformation:

	(1) Emotionality (weighted, R-D)	(2) Emotionality (simple, R-D)	(3) Emotional Valence (R-D)
Misinformation Index	0.082***	0.048***	0.015**
	(0.007)	(0.007)	(0.007)
Other Internet	-0.017	0.040	-0.053**
	(0.025)	(0.025)	(0.025)
Events	-0.069***	-0.026	-0.021
	(0.023)	(0.023)	(0.023)
TV/Radio	-0.035	0.006	-0.035
	(0.022)	(0.022)	(0.022)
Press	-0.074***	0.011	-0.062**
	(0.028)	(0.028)	(0.028)
Constant	0.079**	0.131***	-0.092***
	(0.035)	(0.035)	(0.035)
R^2	0.027	0.012	0.013
Observations	23177	23177	23177

Standard errors in parentheses
Fixed effects included: Year, Season, Author, Reporter
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.4: Heterogeneous Effects of Misinformation on Emotions: Republicans and Democrats

scores assigned to Republicans show greater emotional intensity, while those assigned to Democrats reflect a greater association with content negativity.

To capture the effect of polarizing information, irrespective of the parties considered, table 3.5 reports the regression of the absolute value of the difference for emotionality and emotional valence between Republicans and Democrats on our Misinformation Index. The new outcome variable now expresses the distance of the content from being politically neutral. In other words, the greater the value of the outcome variable, the more polarizing the content. Overall, the results indicate that misinformation does not predict how much polarizing a content is, in terms of its emotional intensity (Column (1) and (2)). However, a significant but small partisan emotional polarization is rather detected in the emotional valence coefficient (column 3), suggesting that more polarizing contents are also characterized by higher emotional negativity.

	(1) Emotionality (weighted, |R-D|)	(2) Emotionality (simple, |R-D|)	(3) Emotional Valence |(R-D|)
Misinformation Index	0.004	-0.005	-0.036***
	(0.007)	(0.007)	(0.007)
Other Internet	-0.016	0.089***	0.088***
	(0.026)	(0.025)	(0.025)
Events	-0.044*	-0.019	-0.006
	(0.024)	(0.023)	(0.023)
TV/Radio	-0.046**	0.018	0.034
	(0.022)	(0.022)	(0.022)
Press	-0.033	-0.025	-0.030
	(0.028)	(0.028)	(0.028)
Constant	-0.017	-0.153***	-0.134***
	(0.035)	(0.035)	(0.035)
R^2	0.005	0.009	0.011
Observations	23177	23177	23177

Standard errors in parentheses
Fixed effects included: Year, Season, Author, Reporter
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.5: Misinformation and Polarization

Finally, table 3.6 shows the results of a SUREG analysis estimating the partisans' differences across specific emotions. The analysis of specific emotions difference shows that Republicans may feel less happiness, but more anger, fear, disgust. Sadness shows a small but significant negative delta, suggesting that the model assigns higher sadness scores to Democrats when they are exposed to misinformative content. These patterns underscore a more negative emotional profile among Republicans compared to Democrats when confronted with misinformation. The highest predictor of partisans' differences is represented by "Happiness", although focusing on the differences in the emotional intensity (Table 3.4) can provide the most comprehensive framework when testing heterogeneity in this context.

	Happiness (R-D)	Sadness (R-D)	Anger (R-D)	Fear (R-D)	Surprise (R-D)	Disgust (R-D)
Misinformation Index	0.042***	-0.015**	0.032***	0.028***	0.017**	0.029***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Other Internet	-0.049*	0.048*	0.037	0.037	-0.007	0.063**
	(0.025)	(0.025)	(0.025)	(0.026)	(0.026)	(0.025)
Events	-0.036	0.026	-0.009	-0.018	-0.036	-0.006
	(0.023)	(0.023)	(0.023)	(0.023)	(0.024)	(0.023)
TV/Radio	-0.042*	0.035	0.009	0.023	-0.002	0.012
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
Press	-0.074***	0.060**	0.027	0.039	-0.033	0.028
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
Constant	-0.065*	0.106***	0.090**	0.112***	0.086**	0.119***
	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)
R^2	0.020	0.009	0.008	0.008	0.004	0.011
Observations	23177					

Standard errors in parentheses
Fixed effects included: year, quarter, frequent authors, frequent reporters
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.6: SUREG Estimates of Emotions Differences (Republican - Democrat)

Lastly, another interesting phenomenon concerns the presence of out-group hostility (opposed to in-group favoritism) when facing news of the opposite party. In this framework, if emotions varies across topics supported by different ideologies, then different emotional responses may point to the presence of out-group animosity when correlated to misinformation. We estimate a model predicting misinformation using proxies for the emotional intensity and valence for both Republicans and Democrats in our specification. Table 3.7 presents results from three model specifications, divided into sub-samples representing the whole sample of news (1), pro-Republicans news (2) and pro-Democrats news (3). The dependent variable is the Misinformation Index, while our specification includes the emotional scores assigned to Democrats (D) and Republicans (R) in the form of emotional valence and intensity (using weighted average computation).

In all specifications, higher positive emotional valence is associated with a lower level of misinformation. Republican-specific valence is consistently negative and statistically significant across all subsamples (column (1): $\beta = -0.058$, $p < 0.01$; column (2): $\beta = -0.021$, $p < 0.05$; column (3): $\beta = -0.104$, $p < 0.01$), but its magnitude is five times larger when observing out-group vs. in-group content (column (2) vs. column (3)). Democrats' emotional valence, on the other hand, is significant only in the pro-Republican news subsample (column (2), $\beta = -0.200$, $p < 0.01$) and in the pooled sample (column (1), $\beta = -0.066$, $p < 0.01$), while it is non-different from zero in the pro-Democratic subsample, suggesting that Democrats' negative emotional responses are overall driven by out-group content. In summary, these substantial differences in the magnitudes of the observed coefficients between Republicans and Democrats result in patterns suggesting that negative emotionality originates from sources aligned with the out-group, which is consistent with theories of selective scrutiny and confirmation bias.

Turning to emotionality, which captures the intensity rather than the valence of a statement, we again observe that greater emotional intensity is associated with lower accuracy. Republican-specific emotionality score has a strong, positive, and significant effect across all models (column (1): $\beta = -0.122$; column (2): $\beta = -0.051$; column (3): $\beta = -0.128$; $p < 0.01$). Democrats' emotional intensity index, in contrast, is significant only in the pro-Republican subsample (column (2): $\beta = 0.038$, $p < 0.01$) and in the full sample, where the coefficient turns negative (column (1): $\beta = -0.025$, $p < 0.01$).

Taken together, these results are consistent with the presence of out-group and in-group biases. Statements perceived as emotionally intense, particularly if negative, are more strongly associated when they come from a political out-group.

	(1) Misinformation Index All	(2) Misinformation Index Pro-Republicans	(3) Misinformation Index Pro-Democrats
Emotional Valence (D)	-0.066***	-0.200***	-0.001
	(0.007)	(0.013)	(0.013)
Emotional Valence (R)	-0.058***	-0.021**	-0.104***
	(0.007)	(0.009)	(0.021)
Emotionality (weighted, D)	-0.025***	0.038***	-0.007
	(0.007)	(0.010)	(0.017)
Emotionality (weighted, R)	0.122***	0.051***	0.128***
	(0.007)	(0.010)	(0.016)
Other Internet	-0.020	0.093***	-0.043
	(0.022)	(0.030)	(0.052)
Events	-0.462***	-0.274***	-0.412***
	(0.020)	(0.027)	(0.045)
TV/Radio	-0.323***	-0.140***	-0.259***
	(0.019)	(0.026)	(0.043)
Press	-0.480***	-0.269***	-0.280***
	(0.024)	(0.035)	(0.056)
Constant	-0.098***	-0.076*	-0.321***
	(0.031)	(0.043)	(0.067)
R^2	0.254	0.218	0.120
Observations	23177	10084	4483

Standard errors in parentheses
Fixed effects included: Year, Season, Author, Reporter
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.7: Out-Group Hostility and In-Group Favoritism

3.3.3 Republican vs. Democrat Content

A final step for our analysis is to verify whether the political slant of statements (pro-Republican vs. pro-Democrats content) affects itself the relationship between misinformation and emotions. Together with the heterogeneous correlations between partisans' emotional scores and misinformation previously shown, the content analysis can contribute to offer an overall overview on how political news are differently associated with their emotional content. We re-estimate our main models by introducing a binary indicator for statements categorized as Pro-Republican, along with an interaction term between the latter and the Misinformation Index. The results, presented in Table 3.8, reveal significant differences in how emotional scores correlate with misinformation depending on the political slant of the content.

In all three columns of Table 3.8, the main effect of the Misinformation Index remains significant, confirming that more inaccurate statements are associated with greater emotionality and more negative valence overall. Moreover, its interaction with the pro-Republican news set is also positive and significant for emotionality (Columns (1) and (2)), and negative for valence (Column (3)). This suggests that the increasing emotionality of misinformation is particularly strong when the misleading content is pro-Republican. In contrast, the main effect of the pro-Republican dummy is negative for emotionality and positive for emotional valence, indicating that accurate pro-Republican content tends to be less emotionally charged and more positively framed than pro-Democratic content.

The emotion-specific results from the SUREG model further confirm such asymmetry (Table 3.9). Inaccurate pro-Republican statements are significantly more likely to evoke high levels of negative emotions such as anger, fear, and disgust, compared to pro-Democratic content. Consistently with the previous findings, accuracy predicts the emotional scores of news, and its interaction with the pro-Republican subsample of news further highlights the presence of heterogeneous effects. For example, the interaction terms for anger, fear, and surprise are all positive and statistically significant, emphasizing that misinformation in pro-Republican content activates stronger negative emotions than pro-Democrat content.

These findings suggest that misinformation is not only emotionally evocative in general but that its emotional content may vary by political alignment, with inaccurate right wing news showing a particularly strong link to emotional language. Together with partisans' heterogeneous emotional responses, we can conclude that there exists a strong link between information and emotion, driven primarily by falsehood of the content, its political topic and readers' political affiliation.

	(1) Emotionality (weighted)	(2) Emotionality (simple)	(3) Emotional Valence
Misinformation Index	0.061***	0.093***	-0.081***
	(0.014)	(0.013)	(0.013)
Pro Republican	-0.095***	-0.179***	0.091***
	(0.018)	(0.017)	(0.016)
Pro Republican × Misinformation Index	0.041**	0.073***	-0.076***
	(0.017)	(0.016)	(0.015)
Other Internet	-0.011	0.005	-0.005
	(0.027)	(0.025)	(0.024)
Events	-0.273***	-0.271***	0.192***
	(0.024)	(0.023)	(0.022)
TV/Radio	-0.311***	-0.218***	0.093***
	(0.023)	(0.022)	(0.020)
Press	-0.243***	-0.144***	0.051*
	(0.031)	(0.029)	(0.028)
Constant	0.033	0.295***	-0.330***
	(0.039)	(0.037)	(0.035)
R^2	0.115	0.135	0.108
Observations	14566	14566	14566

Standard errors in parentheses

Fixed effects included: Year, Season, Author, Reporter

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

Table 3.8: The Effect of Content Accuracy and Political Slant

	Happiness	Sadness	Anger	Fear	Surprise	Disgust
Misinformation Index	-0.068***	0.007	0.104***	0.084***	0.086***	0.109***
	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)	(0.013)
Pro Republican	0.009	-0.286***	-0.169***	0.070***	0.064***	-0.224***
	(0.016)	(0.017)	(0.016)	(0.018)	(0.018)	(0.017)
Pro Republican × Misinformation Index	-0.073***	0.022	0.054***	0.098***	0.081***	0.077***
	(0.015)	(0.016)	(0.016)	(0.017)	(0.018)	(0.016)
Other Internet	0.001	-0.024	0.010	0.060**	-0.032	-0.004
	(0.024)	(0.025)	(0.025)	(0.027)	(0.028)	(0.025)
Events	0.152***	-0.107***	-0.243***	-0.127***	-0.374***	-0.274***
	(0.022)	(0.023)	(0.022)	(0.024)	(0.025)	(0.023)
TV/Radio	0.047**	-0.115***	-0.133***	-0.091***	-0.376***	-0.138***
	(0.020)	(0.021)	(0.021)	(0.023)	(0.023)	(0.021)
Press	0.027	-0.013	-0.094***	-0.048	-0.319***	-0.104***
	(0.027)	(0.029)	(0.028)	(0.031)	(0.031)	(0.029)
Constant	-0.286***	0.144***	0.515***	-0.031	0.045	0.538***
	(0.035)	(0.037)	(0.037)	(0.039)	(0.041)	(0.037)
R^2	0.089	0.052	0.116	0.103	0.176	0.134
Observations	14566					

Standard errors in parentheses

Fixed effects included: year, quarter, frequent authors, frequent reporters

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

Table 3.9: The Effect of Content Accuracy and Political Slant: SUREG

3.4 Conclusion

This paper explored the correlations between emotions and the accuracy of information. Using fact-checked headlines from PolitiFact and tweets collected by the Internet Research Agency, we investigated how emotional intensity and valence relate to misinformation, and how these dynamics vary across political affiliation. Furthermore, we disentangle this association to obtain a comprehensive overview of the role of all six Ekman's emotions in associating with misleading information. Using a large number of observations and a new approach to classify the emotional content of headlines, we provided robust evidence that emotions are deeply intertwined with the falsehood of the content.

Our findings reveal three key insights. First, emotional intensity is positively correlated with inaccuracy, irrespective of its positive or negative representation. Negative emotions such as anger, fear, and disgust emerge as strong predictors of misinformation, while happiness is negatively associated with it. Surprise, which often represents unexpected information, is also strongly associated with falsehoods, suggesting that novelty attracts users' attention. Second, the relationship between emotions and misinformation depends on partisanship, with suggestive evidence of asymmetric effects that reinforce in-group favoritism and out-group hostility. Third, the political content itself, particularly if coming from right-wing sources, further emphasizes the previous results.

Taken together, these results contribute to the existing literature in several ways. We extend previous work by adopting a broader set of emotions and sources, exploiting new text-analysis techniques, and extending the results to observe simulated partisan asymmetries within the same dataset. We provide evidence confirming that the emotional role of misinformation is a pressing concern, highlighting the importance of considering the whole spectrum of emotions, rather than focusing on single related aspects.

From a policy perspective, our results suggest that strategies to fight misinformation should explicitly address its emotional component. Platforms may need to reconsider amplification algorithms that favor outrageous and unexpected content, while interventions that promote awareness could improve users' abilities to discern manipulative content. Importantly, interventions that target emotionally intense news may reduce the spread of misinformation without suppressing political discourse, by limiting the spread of misleading content in presence of signals that flag sources as non-trustworthy.

Future research may consider analyzing the causal impact of emotions on misleading content within the sphere of the demand and supply of (mis)information. From the demand side, it is important to understand which are the mechanisms driving the spread (sharing) of news. Misinformation can circulate more easily on social media because non-attentive users share without reading (McLoughlin et al., 2024). Alternatively, spread and beliefs can depend on how information spreads in groups, later contaminating out-groups. In this scenario, the ability to recall false content makes memory a plausible driver of the demand-side spread. Concerning the supply side, future studies can test the causal impact of how the lack of thematic and ethical constraints allows misinformation producers to craft misleading claims with higher emotional content than producers of accurate information. Building on the mechanisms discussed in

Serra-Garcia (2025), incentives to attract users' attention may lead producers to transmit less complete information, generating biased beliefs among users who do not actively seek more complete information. This process creates a double layer of distortion, originating from the content itself and further amplified by incentives to capture attention.

Although our study consists of a solid methodology, some limitations are to be taken into account. First, we rely on two specific datasets. The PolitiFact and IRA datasets do not cover the universe of misinformation sources. Headlines differ from full length articles and often information is contained in other sources (memes, images, video, AI generated content etc.). Furthermore, because many of the IRA link were containing deleted content, we could not classify the whole sample of news. Second, although we deemed the ability of OpenAI superior to other text analysis algorithms, our emotional measures remain an artificial score, which might be subject to measurement errors. Lastly, our analysis report correlations between emotional content and misinformation, but does not establish causality. However, it serves as a basis for future experiments wanting to detect causal effects among those variables reporting the highest correlations.

References

Acemoglu, D., Ozdaglar, A., and Siderius, J. (2024). A model of online misinformation. *Review of Economic Studies*, 91(6):3117–3150.

Barrera, O., Guriev, S., Henry, E., and Zhuravskaya, E. (2020). Facts, alternative facts, and fact checking in times of post-truth politics. *Journal of public economics*, 182:104123.

Baudouin, J.-Y., Gallian, F., Pinoit, J.-M., and Damon, F. (2025). Arousal, valence, and discrete categories in facial emotion. *Scientific Reports*, 15(1):40268.

Burdea, V. and Woon, J. (2023). Getting it right: Communication, voting, and collective truth finding. Technical report, Discussion Paper.

Ceylan, G., Anderson, I. A., and Wood, W. (2023). Sharing of misinformation is habitual, not just lazy or biased. *Proceedings of the National Academy of Sciences*, 120(4):e2216614120.

Djourelova, M., Durante, R., Motte, E., and Patacchini, E. (2023). Experience, narratives, and climate change beliefs. *Available at SSRN 4680430*.

Fiala, L. and Noussair, C. N. (2017). Charitable giving, emotions, and the default effect. *Economic Inquiry*, 55(4):1792–1812.

Gennaro, G. and Ash, E. (2022). Emotion and reason in political language. *The Economic Journal*, 132(643):1037–1059.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., and Lazer, D. (2019). Fake news on twitter during the 2016 us presidential election. *Science*, 363(6425):374–378.

Henry, E., Zhuravskaya, E., and Guriev, S. (2022). Checking and sharing alt-facts. *American Economic Journal: Economic Policy*, 14(3):55–86.

Kartal, M. and Tyran, J.-R. (2022). Fake news, voter overconfidence, and the quality of democratic choice. *American Economic Review*, 112(10):3367–3397.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., et al. (2018). The science of fake news. *Science*, 359(6380):1094–1096.

Liporace, M. (2021). 4 things nobody tells you about online news.

Martel, C., Pennycook, G., and Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. *Cognitive research: principles and implications*, 5:1–20.

McLoughlin, K. L., Brady, W. J., Goolsbee, A., Kaiser, B., Klonick, K., and Crockett, M. (2024). Misinformation exploits outrage to spread online. *Science*, 386(6725):991–996.

Nyhan, B. and Reifler, J. (2015). The effect of fact-checking on elites: A field experiment on us state legislators. *American journal of political science*, 59(3):628–640.

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., and Petersen, M. B. (2021). Partisan polarization is the primary psychological motivation behind political fake news sharing on twitter. *American political science review*, 115(3):999–1015.

Pennycook, G., Bear, A., Collins, E. T., and Rand, D. G. (2020). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management science*, 66(11):4944–4957.

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., and Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592(7855):590–595.

Pennycook, G. and Rand, D. G. (2019). Fighting misinformation on social media using crowd-sourced judgments of news source quality. *Proceedings of the National Academy of Sciences*, 116(7):2521–2526.

Pfänder, J. and Altay, S. (2025). Spotting false news and doubting true news: a systematic review and meta-analysis of news judgements. *Nature human behaviour*, pages 1–12.

Rasmussen, J., Rathje, S., Van Bavel, J. J., and Robertson, C. (2025). Negativity and identity language have additive effects on online news consumption.

Rathje, S., Roozenbeek, J., Van Bavel, J. J., and Van Der Linden, S. (2023). Accuracy and social motivations shape judgements of (mis) information. *Nature Human Behaviour*, 7(6):892–903.

Rathje, S., Van Bavel, J. J., and Van Der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the national academy of sciences*, 118(26):e2024292118.

Serra-Garcia, M. (2025). The attention–information tradeoff. Technical report, CESifo Working Paper.

Serra-Garcia, M. and Gneezy, U. (2021). Mistakes, overconfidence, and the effect of sharing on detecting lies. *American Economic Review*, 111(10):3160–3183.

Soroka, S., Fournier, P., and Nir, L. (2019). Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences*, 116(38):18888–18892.

Sultan, M., Tump, A. N., Ehmann, N., Lorenz-Spreen, P., Hertwig, R., Gollwitzer, A., and Kurvers, R. H. (2024). Susceptibility to online misinformation: A systematic meta-analysis of demographic and psychological factors. *Proceedings of the National Academy of Sciences*, 121(47):e2409329121.

Thaler, M. (2024). The fake news effect: Experimentally identifying motivated reasoning using trust in news. *American Economic Journal: Microeconomics*, 16(2):1–38.

Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *science*, 359(6380):1146–1151.

Appendix A

A.1 Emotional Intensity and Valence among Simulated Partisans

In this appendix, we report the complete estimation results for the additional models introduced in Table 3.2. These models extend the main analysis by computing emotional intensity and valence from the perspective of readers with different partisan affiliations (Republicans and Democrats).

	(1) Emotionality (weighted, R)	(2) Emotionality (simple, R)	(3) Emotional Valence (R)
Misinformation Index	0.143***	0.154***	-0.118***
	(0.007)	(0.007)	(0.007)
Other Internet	0.043*	0.054**	-0.019
	(0.025)	(0.025)	(0.025)
Events	-0.233***	-0.164***	0.121***
	(0.023)	(0.023)	(0.023)
TV/Radio	-0.211***	-0.060***	0.016
	(0.022)	(0.022)	(0.022)
Press	-0.218***	-0.058**	0.006
	(0.027)	(0.028)	(0.028)
Constant	0.032	-0.011	-0.012
	(0.034)	(0.034)	(0.035)
R^2	0.086	0.068	0.049
Observations	23178	23178	23178

Standard errors in parentheses
Fixed effects included: Year, Season, Author, Reporter
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A1: OLS Emotionality & Valence - Republicans

	(1) Emotionality (weighted, D)	(2) Emotionality (simple, D)	(3) Emotional Valence (D)
Misinformation Index	0.059***	0.110***	-0.137***
	(0.007)	(0.007)	(0.007)
Other Internet	0.059**	0.019	0.030
	(0.025)	(0.025)	(0.025)
Events	-0.160***	-0.138***	0.145***
	(0.023)	(0.023)	(0.023)
TV/Radio	-0.173***	-0.064***	0.049**
	(0.022)	(0.022)	(0.022)
Press	-0.141***	-0.066**	0.065**
	(0.028)	(0.028)	(0.028)
Constant	-0.047	-0.119***	0.073**
	(0.035)	(0.034)	(0.034)
R^2	0.055	0.058	0.064
Observations	23179	23179	23179

Standard errors in parentheses
Fixed effects included: Year, Season, Author, Reporter
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A2: OLS Emotionality & Valence - Democrats

	Happiness (R)	Sadness (R)	Anger (R)	Fear (R)	Surprise (R)	Disgust (R)
Misinformation Index	-0.103***	0.013*	0.133***	0.148***	0.180***	0.154***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Other Internet	0.006	0.035	0.042*	0.050**	-0.018	0.043*
	(0.025)	(0.025)	(0.025)	(0.025)	(0.024)	(0.025)
Events	0.121***	-0.019	-0.132***	-0.070***	-0.349***	-0.159***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.022)	(0.023)
TV/Radio	0.028	-0.011	0.015	0.003	-0.317***	-0.011
	(0.022)	(0.022)	(0.022)	(0.022)	(0.021)	(0.022)
Press	0.015	0.079***	-0.009	0.004	-0.310***	-0.044
	(0.028)	(0.028)	(0.028)	(0.028)	(0.026)	(0.028)
Constant	-0.039	-0.060*	0.004	-0.072**	0.136***	0.032
	(0.035)	(0.035)	(0.035)	(0.034)	(0.033)	(0.034)
R^2	0.044	0.020	0.052	0.060	0.150	0.061
Observations	23178					

Standard errors in parentheses

Fixed effects included: year, quarter, frequent authors, frequent reporters

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

Table A3: SUREG Specific Emotions - Republicans

	Happiness (D)	Sadness (D)	Anger (D)	Fear (D)	Surprise (D)	Disgust (D)
Misinformation Index	-0.153***	0.024***	0.105***	0.126***	0.167***	0.130***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Other Internet	0.056**	-0.003	0.009	0.020	-0.013	-0.015
	(0.025)	(0.025)	(0.025)	(0.025)	(0.024)	(0.025)
Events	0.167***	-0.038	-0.126***	-0.056**	-0.322***	-0.156***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.022)	(0.023)
TV/Radio	0.073***	-0.038*	0.007	-0.015	-0.317***	-0.022
	(0.022)	(0.022)	(0.022)	(0.022)	(0.021)	(0.022)
Press	0.091***	0.029	-0.033	-0.027	-0.285***	-0.071**
	(0.027)	(0.028)	(0.028)	(0.028)	(0.026)	(0.028)
Constant	0.024	-0.137***	-0.078**	-0.161***	0.068**	-0.077**
	(0.034)	(0.035)	(0.035)	(0.034)	(0.032)	(0.034)
R^2	0.072	0.020	0.046	0.063	0.156	0.058
Observations	23179					

Standard errors in parentheses

Fixed effects included: year, quarter, frequent authors, frequent reporters

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

Table A4: SUREG Specific Emotions - Democrats

A.2 Robustness: Explanatory Power and the IRA dataset

Here we report the replication of the main model (Table 3.2) omitting the Misinformation Index for drawing conclusions about R-squared comparison. In addition, we present the robustness analysis using the data collected by the Internet Research Agency (IRA).

	(1) Emotionality (weighted)	(2) Emotionality (simple)	(3) Emotional Valence
Other Internet	-0.010	0.013	-0.007
	(0.024)	(0.023)	(0.022)
TV/Radio	-0.377***	-0.233***	0.082***
	(0.021)	(0.020)	(0.019)
Press	-0.337***	-0.235***	0.118***
	(0.026)	(0.025)	(0.024)
Events	-0.363***	-0.321***	0.211***
	(0.022)	(0.021)	(0.020)
Constant	-0.138***	-0.065**	-0.072**
	(0.033)	(0.032)	(0.030)
R^2	0.117	0.091	0.056
Observations	23338	23338	23338

Standard errors in parentheses

Fixed effects included: Year, Season, Author, Reporter

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

Table A5: OLS Emotionality & Valence: R-squared Comparison

To assess the accuracy of the content in the IRA dataset, we follow the procedures adopted by McLoughlin et al. (2024). Starting from raw data, we expanded the URL associated with the content of the tweet, isolating and matching the domain with the data provided by McLoughlin et al. (2024), which classifies the domain as either "Trustworthy" or "Misinformation". We collected the emotional score of the tweet content's headlines and analyzed the impact of a binary variable for Misinformation (False) on the emotionality and valence of the news content. We estimate the same set of regression models previously discussed, including in the specification the binary variable False, a dummy "IRA" to capture the datasets differences and its interaction with the misinformation dummy.

The results are presented in Table A6. Columns (1) and (2) show that false information in the Politifact sample is associated with significantly higher emotionality across both weighted and unweighted measures. Column (3) confirms that misinformation is also linked to a more negative emotional valence. Importantly, the two dataset are not directly comparable, as the Politifact sample embed information from a broader set of sources, while the IRA sample contains only posts published on Twitter (today named after X). Here, we observe that the accurate IRA content is generally more emotional and reports more positive valence on average. Yet, the interaction term between False and IRA is negative and statistically significant for valence, implying that false content from IRA sources tends to be even more negatively emotional.

In Tables A7 and A8, we report the univariate regressions between our measures of the emotional content and a dummy equal to 1 if the statement is classified as false in the Politifact (Table A7) and IRA (Table A8) samples. The magnitude of the coefficient in the Politifact

	(1) Emotionality (weighted)	(2) Emotionality (simple)	(3) Emotional Valence
False	0.405***	0.454***	-0.342***
	(0.016)	(0.016)	(0.015)
IRA	0.490***	0.166***	0.184***
	(0.015)	(0.016)	(0.015)
False × IRA	-0.068***	0.026	-0.074***
	(0.024)	(0.024)	(0.024)
Constant	-0.391***	-0.283***	0.079***
	(0.013)	(0.013)	(0.013)
R^2	0.043	0.040	0.055
Observations	38480	38480	38480

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Robustness: Emotionality and Valence in IRA vs. Politifact

sample differs from that shown when using the Misinformation Index (Table 3.2) due to two main reasons: first, the exclusion of the sources as a control, and second, the conversion of the Misinformation Index into a dummy variable. Overall, the results are in line to those found when using the IRA dataset (Table A8).

	(1) Emotionality (weighted)	(2) Emotionality (simple)	(3) Emotional Valence
False	0.405***	0.454***	-0.343***
	(0.016)	(0.015)	(0.013)
R^2	0.033	0.046	0.033
Observations	19762	19762	19762

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Emotionality and Valence: Politifact

	(1) Emotionality (weighted)	(2) Emotionality (simple)	(3) Emotional Valence
False	0.337***	0.480***	-0.416***
	(0.018)	(0.019)	(0.020)
R^2	0.019	0.033	0.022
Observations	18718	18718	18718

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A8: Emotionality and Valence: IRA

Finally, we replicate our emotion-specific SUREG analysis to decompose the relationship between misinformation and specific emotions. As shown in Table A9, false statements from Politifact are once again consistently associated with higher levels of anger, fear, surprise, and disgust, and significantly lower happiness. Interestingly, the interaction terms reveal that IRA false statements show amplified expressions of anger and disgust, while the effects on fear and surprise are attenuated compared to non-IRA sources.

	Happiness	Sadness	Anger	Fear	Surprise	Disgust
False	-0.274***	0.060***	0.387***	0.479***	0.548***	0.422***
	(0.015)	(0.016)	(0.015)	(0.016)	(0.016)	(0.015)
IRA	0.323***	0.147***	-0.162***	0.188***	0.272***	-0.145***
	(0.015)	(0.016)	(0.015)	(0.016)	(0.015)	(0.015)
False × IRA	-0.079***	-0.048**	0.160***	-0.137***	-0.062***	0.211***
	(0.023)	(0.024)	(0.023)	(0.024)	(0.024)	(0.023)
Constant	-0.016	-0.088***	-0.122***	-0.290***	-0.348***	-0.150***
	(0.013)	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)
R^2	0.067	0.003	0.066	0.034	0.049	0.075
Observations	38480					

Standard errors in parentheses

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

Table A9: Robustness: SUREG: IRA vs. Politifact

Appendix B

Table B1 reports the scores assigned using OpenAI's API for a subsample of news included in our analysis. The selection of this subsample was made from a pilot dataset, which include the request of providing a reasoning behind the scores assignment. For each news' ID, Table B2 reports the explanation provided. Overall, we deemed OpenAI capable of motivating its answers, without detecting anomalies in the classification exercise.

ID	News	Happiness	Sadness	Anger	Fear	Surprise	Disgust
1	2022 was the biggest tourism year ever in West Virginia	6	1	1	1	4	1
2	Video shows a family learning their debt was resolved by the American Debt Relief program for American citizens and legal residents who have over $20,000 in credit card debt.	6	2	1	2	5	1
3	Lego donates model MRI kits to hospitals to help children understand the procedure and reduce their anxiety.	6	2	1	2	4	1
4	Rise in U.S. infant mortality rate connected to COVID-19 vaccines.	1	6	5	6	5	5
5	Elon Musk is dead.	1	6	3	5	6	2
6	Video shows Israel Zionist blowing up people in hospitals in 2023.	1	6	7	6	5	7
7	Sen. Dianne Feinstein tweeted about getting COVID-19 and flu vaccines days before her death.	2	6	3	4	5	2
8	Ground zero in New York - I remember standing there the next day.	1	6	4	5	3	4
9	The WHO is developing a pandemic treaty that will remove current human rights protections, enforce surveillance and censorship and eliminate freedom of speech.	1	5	6	7	6	6
10	Sen. Rick Scott wants to . end Social Security and Medicare coverage.	1	5	6	6	5	5
11	FEMA and Red Cross are confiscating donations meant for Hawaii wildfire victims.	1	5	6	4	5	6
12	The reality is more people are dying of bad climate change policies than they are of actual climate change.	1	5	6	5	4	6
13	Women vaccinated against COVID-19 face unknown side effects that could perpetually go down through the lineage to their children.	1	4	3	6	5	3
14	The federal government has just declared a major food shortage emergency in six states.	1	5	4	6	5	3
15	Photos are proof that Hawaii fires were caused by a direct energy weapon assault on the people.	1	4	5	6	5	5
16	Official statement - Putin assassinated in the Kremlin!	2	4	5	6	7	5
17	Hiroshima and Nagasaki were never nuked.	1	5	4	3	6	4
18	Fox News reported on an energy-saving device that can save 90% of electricity from Elon Musk.	5	1	1	2	6	1
19	U.S. Navy SEALs rescued 807 children who were being kept as sex slaves on a barge in San Francisco Bay.	3	6	6	5	5	7
20	The United Nations has warned Christians that if we do not fully accept the legalization of pedophilia that we will be excluded from participating in society.	1	5	6	7	6	7

Table B1: Emotions Scores Example

Table B2: Reasoning Behind Score Assignment

ID	News	Reasoning
1	2022 was the biggest tourism year ever in West Virginia	The headline 2022 was the biggest tourism year ever in West Virginia primarily conveys a positive sentiment, suggesting success and growth, which is why happiness is rated high. There is little to no indication of sadness, anger, fear, or disgust, as the headline does not mention any negative aspects or consequences. The element of surprise is moderately present, as the achievement of a record-breaking year could be unexpected or noteworthy.
2	Video shows a family learning their debt was resolved by the American Debt Relief program for American citizens and legal residents who have over $20,000 in credit card debt.	The headline likely evokes a strong sense of happiness as it describes a family having their debt resolved, which is typically a positive and relieving outcome. There is also a significant element of surprise, as having a large debt resolved unexpectedly can be quite astonishing. Sadness, anger, fear, and disgust are likely low because the resolution of debt is generally seen as a positive event, reducing negative emotions.
3	Lego donates model MRI kits to hospitals to help children understand the procedure and reduce their anxiety.	The headline conveys a positive and supportive action by Lego, which is likely to evoke happiness due to the charitable and helpful nature of the donation. The sadness and fear scores are low because the initiative aims to reduce anxiety and make a potentially intimidating medical procedure more understandable for children. Anger and disgust are minimal as there is no negative or offensive aspect in the headline. The surprise score is moderate, as the idea of using Lego models in a medical context is somewhat unexpected and innovative.
4	Rise in U.S. infant mortality rate connected to COVID-19 vaccines.	The headline suggests a negative and concerning situation, which is likely to evoke high levels of sadness and fear due to the mention of infant mortality. Anger and disgust are also significant because the headline implies a controversial connection between COVID-19 vaccines and infant deaths, which can provoke strong emotional reactions. Surprise is moderately high as the claim may be unexpected to many readers. Happiness is very low as the content of the headline is distressing.
5	Elon Musk is dead.	The headline Elon Musk is dead is likely to evoke high levels of sadness and surprise, as the death of a prominent and influential figure like Elon Musk would be unexpected and impactful. Fear may also be significant due to potential implications for industries and markets he influences. Anger could arise from those who feel strongly about his work or persona, while happiness is expected to be very low, as death is generally not associated with positive emotions. Disgust is likely minimal but could be present depending on individual perspectives on Musk.
6	Video shows Israel Zionist blowing up people in hospitals in 2023.	The headline evokes strong negative emotions due to its content. The mention of blowing up people in hospitals suggests violence and harm, which typically leads to high levels of sadness, anger, and disgust. Fear is also significant as the event described is threatening and distressing. Surprise is moderately high because such events are shocking and unexpected. Happiness is at its lowest because the headline describes a tragic and violent situation.
7	Sen. Dianne Feinstein tweeted about getting COVID-19 and flu vaccines days before her death.	The headline primarily evokes sadness due to the mention of Sen. Dianne Feinstein's death. There is an element of surprise because the headline notes her recent activity on social media just days before her passing, which can be unexpected for readers. Fear is moderately present as it touches on health concerns related to COVID-19 and flu, which are ongoing public health issues. Anger and disgust are less prominent but may arise from the context of a public figure's death and the ongoing discourse around vaccines. Happiness is minimal as the overall tone of the headline is somber.

Continued on next page

ID	News	Reasoning
8	Ground zero in New York - I remember standing there the next day.	The headline refers to Ground zero in New York, which is commonly associated with the aftermath of the September 11 attacks. This context evokes strong emotions of sadness and fear due to the tragic loss of life and the impact of the event. Anger and disgust are also present, reflecting the feelings towards the attack itself. Surprise is moderate, as the event was unexpected, but the reference to the next day suggests a period of reflection rather than immediate shock. Happiness is rated very low, as the context is somber and reflective of a tragic event.
9	The WHO is developing a pandemic treaty that will remove current human rights protections, enforce surveillance and censorship and eliminate freedom of speech.	The headline suggests a significant and negative change to fundamental human rights, which is likely to evoke strong emotions. The mention of removing human rights protections and eliminating freedom of speech can lead to high levels of fear, anger, and disgust due to the perceived threat to personal freedoms and autonomy. Surprise is also high, as such drastic measures are unexpected. Sadness is present due to the potential loss of rights, while happiness is very low given the negative implications of the headline.
10	Sen. Rick Scott wants to . end Social Security and Medicare coverage.	The headline suggests a significant policy change that could negatively impact many people, leading to high levels of anger and fear due to the potential loss of essential benefits. Sadness is also present as the headline implies a loss of security for those relying on these programs. Surprise is moderately high because such a proposal can be unexpected and shocking to the public. Disgust is also present due to the perceived insensitivity of the proposal. Happiness is very low as the headline does not convey any positive outcomes.
11	FEMA and Red Cross are confiscating donations meant for Hawaii wildfire victims.	The headline suggests a negative situation where donations intended for wildfire victims are being confiscated by FEMA and the Red Cross. This implies a sense of injustice and wrongdoing, which likely evokes high levels of anger and disgust. The term confiscating can also lead to surprise, as it is unexpected for organizations meant to help to be involved in such actions. Sadness is present due to the plight of the victims who are not receiving the aid they need. Fear is moderately present as there may be concerns about the effectiveness and trustworthiness of these organizations in crisis situations. Happiness is very low as the overall tone of the headline is negative.
12	The reality is more people are dying of bad climate change policies than they are of actual climate change.	The headline suggests a critical and negative perspective on climate change policies, implying that these policies are causing more harm than the climate change itself. This evokes strong emotions of anger and disgust, as it suggests a failure or mismanagement that leads to loss of life. Sadness is also present due to the mention of people dying. Fear is moderately present as the headline implies a threat to human life. Surprise is somewhat present as the statement challenges common perceptions about climate change. Happiness is absent as the content is inherently negative.
13	Women vaccinated against COVID-19 face unknown side effects that could perpetually go down through the lineage to their children.	The headline suggests a potential threat to future generations due to unknown side effects of COVID-19 vaccination, which can evoke high levels of fear and surprise. The mention of unknown side effects and perpetually go down through the lineage can cause anxiety and concern, leading to a high fear score. The surprise score is also elevated due to the unexpected nature of the claim. Sadness is moderately present as the idea of negative impacts on children can evoke a sense of loss or concern. Anger and disgust are present but to a lesser extent, as the headline might provoke frustration or aversion towards the situation or the entities involved. Happiness is rated very low as the headline does not convey any positive emotions.

Continued on next page

ID	News	Reasoning
14	The federal government has just declared a major food shortage emergency in six states.	The headline conveys a serious and alarming situation, which is likely to evoke high levels of fear due to the implications of a food shortage emergency. Sadness is also significant as it suggests potential suffering and hardship for those affected. Anger may arise from frustration or perceived mismanagement leading to such a crisis. Surprise is moderately high, as a major food shortage is an unexpected and significant event. Disgust is present but less pronounced, possibly due to dissatisfaction with the situation. Happiness is minimal, as the headline does not contain any positive elements.
15	Photos are proof that Hawaii fires were caused by a direct energy weapon assault on the people.	The headline suggests a conspiracy theory involving a direct energy weapon assault, which can evoke strong negative emotions. Fear is likely high due to the implication of a deliberate attack, which can be alarming. Anger and disgust are also elevated, as the idea of an assault on people can provoke outrage and revulsion. Sadness is present due to the context of fires in Hawaii, which suggests destruction and loss. Surprise is moderately high because the claim of a direct energy weapon is unexpected and unusual. Happiness is very low, as the overall tone of the headline is negative and distressing.
16	Official statement - Putin assassinated in the Kremlin!	The headline Official statement - Putin assassinated in the Kremlin! is likely to evoke a high level of surprise due to the unexpected nature of an assassination of a prominent world leader. Fear is also high, as such an event could lead to political instability and uncertainty. Anger and disgust are moderately high, as assassination is a violent act that can provoke strong negative reactions. Sadness is present, as the loss of life and potential consequences can be distressing. Happiness is low, as the context of assassination generally does not evoke positive emotions.
17	Hiroshima and Nagasaki were never nuked.	The headline Hiroshima and Nagasaki were never nuked is likely to evoke surprise due to its controversial and historically inaccurate claim, hence the high score for surprise. Sadness and anger are also significant because the statement dismisses the tragic events and suffering experienced by the victims, which can be distressing and infuriating to those aware of the historical facts. Disgust is present due to the insensitivity of denying a well-documented historical atrocity. Fear is moderate, as such statements can contribute to misinformation and historical revisionism, which can be concerning. Happiness is very low, as the headline does not convey any positive or joyful sentiment.
18	Fox News reported on an energy-saving device that can save 90% of electricity from Elon Musk.	The headline about an energy-saving device that can save 90% of electricity is likely to evoke a high level of surprise due to the significant claim of energy savings. Happiness is also moderately high, as such a device could have positive implications for energy efficiency and cost savings. Sadness, anger, and disgust are low because the headline does not present negative or distressing information. Fear is slightly present, possibly due to skepticism or concern about the feasibility or implications of such a device.
19	U.S. Navy SEALs rescued 807 children who were being kept as sex slaves on a barge in San Francisco Bay.	The headline evokes a strong sense of disgust due to the horrific nature of the crime described. Sadness and anger are also high, as the situation involves the exploitation and suffering of children. Fear is present due to the alarming nature of such crimes occurring. Surprise is moderately high, as the scale and location of the incident are unexpected. Happiness is relatively low but present due to the successful rescue operation by the U.S. Navy SEALs.
20	The United Nations has warned Christians that if we do not fully accept the legalization of pedophilia that we will be excluded from participating in society.	The headline is likely to evoke strong negative emotions. The mention of legalization of pedophilia is highly likely to cause disgust and anger due to the moral and ethical implications. Fear is also a significant factor, as the idea of being excluded from participating in society can be alarming. Surprise is present due to the unexpected and shocking nature of the claim. Sadness may arise from the perceived threat to societal values. Happiness is rated very low as the content of the headline is distressing and controversial.

Appendix C

In Figure C1, we present the Python script used to automate the emotion analysis of textual news through the OpenAI API. The script reads an input dataset and, for each headline, prompts the model to assign values on a 1–7 scale for six basic emotions: “happiness”, “sadness”, “anger”, “fear”, “surprise”, and “disgust”. The model returns the results in a structured JSON format, where each emotion serves as a key paired with its corresponding numerical score (e.g., {”happiness”: 5, ”sadness”: 2, ...}).

A temperature value of 0 is used to ensure deterministic and replicable outputs, minimizing randomness in the model’s responses. Once completed the classification exercise, the JSON outputs are parsed and divided into separate emotion-specific columns within the dataset. The final dataframe is then exported to an Excel file for further statistical analysis. This automated procedure ensures standardized, consistent, and scalable emotional scoring across all textual inputs.

```python
import pandas as pd
from openai import OpenAI
import logging

client = OpenAI(api_key='my_API_key')
df = pd.read_excel('input_file_path')
emotions = ["happiness", "sadness", "anger", "fear", "surprise", "disgust"]

def analyze_emotion(Statement):
    prompt = (
        f"Analyze the statement and assign values (1-7) for each emotion:\n\n"
        f"Statement: {Statement}\n\n"
        f"Return JSON with emotions as keys and scores as values."
    )
    try:
        response = client.chat.completions.create(
            model="gpt-4o",
            messages=[
                {"role": "system",
                 "content": "You are a psychological expert. Return emotion values in JSON."},
                {"role": "user", "content": prompt}
            ],
            temperature=0
        )
        raw_response = response.choices[0].message.content.strip()
        if "```" in raw_response:
            raw_response = raw_response.replace("```json", "").replace("```", "").strip()
        emotion_scores = eval(raw_response)
        return emotion_scores
    except Exception as e:
        logging.error(f"API call failed for '{Statement}': {e}")
        return {emotion: 0 for emotion in emotions}

df['emotion_scores'] = df['Statement'].apply(analyze_emotion)
for emotion in emotions:
    df[emotion] = df['emotion_scores'].apply(lambda x: x.get(emotion, None))
df.drop(columns=['emotion_scores'], inplace=True)
df.to_excel('output_file_path', index=False)
```

Figure C1: Python script for automated emotion analysis using the OpenAI API