The Impact of Audience Size on Image Concerns: Evidence from a Charity Dictator Game

Sem Manna

January 11, 2024

Abstract: Can the mere presence of noninteractive observers motivate prosocial behavior? Are audience effects monotonic in the number of onlookers? The empirical literature provides mixed findings to these questions. I address them through clean experimental evidence from a modified dictator game involving an external, noninteractive audience of variable size and a charity receiver. The simplicity of this design allows to isolate audience effects from confounding features of experimental designs. I find that the presence of an audience increases donations by 24%, with every audience member increasing donations for an estimated average of 5%. I also outline a parsimonious theoretical framework for size-dependent image concerns that lends itself to natural extensions across various social contexts where actions can be interpreted as signals over private traits.

Acknowledgments

I am most grateful to Alessandro Stringhi for constant feedback and guidance in the development of this draft and for making this project possible. I am also grateful to participants at the 11^{th} Meeting of the Behavioral and Experimental Economics Network (BEEN) at Bologna and to Prof. Luca Braghieri, Nicola Campigotto, and Damiano Paoli for helpful suggestions. I am also indebted to my supervisor, Prof. Battigalli especially for advice on a theoretical section which is still a work in progress. I also want to thank lab managers Marta Maxia and Massimiliano Pozzi, along with the staff at the Bocconi Experimental Laboratory for the Social Sciences, for their support and professionalism.
...

Contents

1 Introduction

Do we behave more morally when we are being observed? The impact of observation on behaviour is one of the earliest questions in philosophical and then scientific inquiry. In Plato's *Republic*, Glaucon tells Socrates the tale of the Ring of Gyges which, by granting invisibility to those who wear it, challenges the morality of men by allowing them to defy laws and social norms whilst avoiding punishment and bad reputations. Socrates replies that the virtuous man is him who chooses not to wear the ring, spelling out the glaring intuition that moral behaviour should stem from intrinsic motives, rather than image concerns. Yet, most of us are not so virtuous and Glaucon's insight that observability induces prosocial behaviour seems vindicated by empirical evidence: when more socially exposed, individuals tend to donate more to charities (see Glazer and Konrad 1996, Cañigueral and Hamilton 2019) and churches (Soetevent 2005), work harder for public goods (Filiz-Ozbay and Ozbay 2014), become more productive (Falk and Ichino 2006) and trustworthy in the eyes of others (Tadelis 2011), are less likely to lie (Fischbacher and Föllmi-Heusi 2013, Gneezy et al. 2018, Dufwenberg and Dufwenberg 2018, and Abeler et al. 2019), and are more likely to turn out to vote (Gerber et al. 2008 and DellaVigna et al. 2016) and abide by a social norm (Schram et al. 2013).

These effects of observability on prosocial behavior could stem from fear of punishment and retaliation, loss of social value as a partner for cooperative endeavors, or simply from an innate preference for being perceived as "good", irrespective of potential future interactions. If a good image is valuable per se, promoting observability could help incentivize prosocial behavior in a range of social contexts where people have one-off quasi-anonymous interactions or where norms are weakly enforced and free-riding is widespread.

Understanding whether passive observation from noninteractive audiences can induce altruism and how the number of observers shapes this influence then becomes crucial to learn about the role of moralistic emotions in decision-making and to assess the evolutionary purpose they may have once served. For instance, altruism can be interpreted as a costly signal of one's fitness as a partner for reciprocal cooperation (see Trivers (1971) and Delton et al. (2011) among others). Moreover, in light of the deterioration of public discourse online and following the social media-related increase in the size of audiences, our findings on how observers can incentivize prosocial behavior even in an anonymous setting become especially relevant.

In this thesis, I present a simple theoretical model which motivates a laboratory experiment to unequivocally investigate the following questions:[1]

1. *Can observation from an anonymous, external, and noninteractive audience motivate prosocial behavior?*
2. *Are audience effects monotonic in the number of onlookers?*

My results suggest that the presence of an anonymous, external, and noninteractive audience is sufficient to increase the share of individuals choosing to split their endowment with a charitable organizations by 24%. Even for small audiences (of 2 or 7 people) the intensity of this effect seems to depend on the number of onlookers, with larger audiences yielding a greater increase in donations. On average, every additional audience member is associated with a 5% increase in the share of equal splits compared to the baseline with no audience.

These questions relate to a longstanding yet ongoing academic inquiry on social image (see section 3.4 of Fehr and Charness (2023) for a summary).[2] Moreover, the extant literature offers conflicting answers, often polluted by confounding features of the experimental design employed (such as allowing subjects to communicate) or by the use of ambiguous forms of social influence involving active audiences or repeated games, which make observers payoff-relevant in the strategic analysis (for a meta-analysis of the literature on observability on prosociality, see Bradley et al. (2018)). Because they introduce observability in different ways, those studies conflate different social phenomena under the same terminology, blurring relevant nuances in the complexity of human interactions.[3] To offer reliable findings, I try to minimize many of these confounders by pairing clean definitions of audience effects and image concerns with a theory-driven experimental design involving a minimal and neat manipulation. To the best of my knowledge, this is the first empirical investigation of size-dependent image concerns.

Observability and Audience Effects Since its early days, the experimental evidence on social facilitation showed that not all observers are alike.[4] Underestimating how different audiences –

[1]The experiment, including hypotheses, design, sampling and analysis plan, had been pre-registered on OSF at: https://osf.io/yfhdp.

[2]Indeed, the behavioural effects of a social presence were the subject of one of the very first experiments in Social Psychology in 1898, when Norman Triplett showed that children's performance in a bike race improved when competing against a counterpart as opposed to racing against the clock (Triplett 1898).

[3]Albeit often overlooked in the Economics literature, some categorizations of the various forms of observability are well established. Zajonc (1965) distinguished *social facilitation* between *audience effects*, stemming from the presence of passive spectators, and *co-action effects*, arising from the presence of individuals engaged in the same activity. The latter are commonly referred to as *peer effects* in the Economics literature.

[4]For instance, Cottrell et al. (1968) studied student's performance in a verbal task and found that a watching audience has a stronger influence than the mere presence of others who are neither co-actors nor observers, highlighting the impact of observers as attentive, watching agents as opposed to the mere presence of non-watching individuals.

including those formed by peers, strategically relevant coplayers, or active, communicative audiences – can impact on how observability can induce prosociality, may introduce both theoretical and methodological confounders (Bradley et al. 2018).

In particular, if the subject of scientific inquiry are image concerns, more active and communicative audiences may confound results through verbal or nonverbal cues which can influence beliefs and expectations which elicit other valanced emotions such as guilt (Attanasi et al. 2019) or trust (Charness and Dufwenberg 2006). An audience of coplayers may instead produce *peer effects* (Falk and Ichino 2006) or induce players to behave as to indirectly influence the behaviour of others – for instance by trying to comply to or establish a social norm (De Cremer and Barker (2003), Schram et al. (2013)) or through reciprocity (Rabin 1993). Publicly revealing actions to all participants, as opposed to being individually observed by an external audience, may further confound results as people tend to conform (both positively and negatively) to the most common or socially appropriate conduct when publicly observed (Andersson et al. 2020). These effects may differ from "pure" social image concerns about unobservable (private) traits. Yet, manipulations studying the link between observability and altruism often fail to discern the two. Similarly, assigning monitoring roles to players may shape norms and expectations (as in Izuma et al. 2011) possibly inducing experimenter-demand effects by allowing subjects to presume what's being studied or subtly prescribing some behaviour (Zizzo 2010).

Finally, the context in which social interaction occurs, along with features of the observing audience or of the interactive task at hand, could influence aspects of the strategic interactions that have been shown to be theoretically or empirically relevant, such as the reference group to which individuals strive to conform with (Bernheim 1994), their normative or empirical expectations over the social norm in place (d'Adda et al. 2020) or the type of social image which they value (Moisan et al. 2023). Most notably, in their experiment Moisan et al. (2023) found that graduate students playing a public good game contributed significantly *less* when their actions and photos were shared with future participants from younger cohorts. The authors still attribute this behaviour to image concerns, positing that students might wish to be perceived as "smart and successful" (an alternative explanation could be that they want to look tough). In this case, audience characteristics (i.e., age) or the social environment may shape what students want to signal. Possibly, individuals manage their reputation in the eyes of different audiences strategically (Emler 1990).

As in some of the cases outlined, experiments studying the effects of observability on prosocial behaviour often focus on different forms of social influence and reach surprising or apparently conflicting results (Bradley et al. 2018). For these reasons, the experiment employed here involves a simple task with clear social valence (donating to a charity) and employs an anonymous and external audience, of which active players only know the size.

In order to minimize the risk of misattributing the results of the experiment, I then begin by borrowing Hamilton and Lind (2016)'s stark definition of *audience effects* as:

> [behavioural changes] "caused by being observed by another person, or [by] the belief that one is being observed [...] This requires [...] some level of awareness that the other is watching, that is, awareness of the perceptual state of the other".

Accordingly, in the experimental design I try to cleanly induce observability in the most unconfounded way possible by simply sharing terminal information about actions after they have been played, along with indirect identity information such as IDs (thus inducing "pseudo-observability"), to an audience of anonymous, external agents without financial stakes in the outcome of the game.

By inducing observability in this way to study social image concerns, this thesis is related to Regner (2021), whose experiment had subjects play a dictator game while a third player observes the outcome. This work significantly differs from his in relevant methodological and theoretical ways. First, in this experiment roles are preassigned and fixed, while in Regner (2021) players rotate across all roles in 3 stages of the same game. While transfers where announced at the end of the experiment, making subject play multiple roles may still confound the results. Moreover, as argued, an audience of coplayers may differ from an external one. Second, instead of a standard Dictator Game, I employ a *Charity Dictator mini-Game* where the choice is dichotomous and the (pseudo) co-player is a charity organization. Allowing Dictators to choose only between an unambiguously selfish and an altruistic allocation allows us to reduce the uncertainty over the social valence of the actions, while charity co-players mitigate concerns over a competitive contexts where agents may wish to signal other traits (as in Moisan et al. (2023)) or feel excused to behave selfishly (Dana et al. 2007). Our result also differ, as I find significant effects for observability even in the absence of public identification, while their result on this point is only qualitative. Finally, I expand from their and related literature by studying how those effects may vary with the number of observers.

Image Concerns I follow Battigalli and Dufwenberg (2022) by neatly defining *image concerns* as preferences over others' ex-post beliefs about private *good or bad traits* which are imperfectly observed. The main behavioural implication of image concerns (or, equivalently, *intrinsic reputation concerns*) is that agents may choose to forgo material payoffs to influence others' opinion about them, even absent strategic implications in terms of future material payoffs (Battigalli et al. 2019). This is consistent with theories of impure altruism (Andreoni 1990), image signalling (Bénabou and Tirole 2006), social conformity (Bernheim 1994), and lying-aversion (Gneezy et al. 2018).

The theoretical model presented here remains general, simply extending and discussing the utility formulation from Battigalli and Dufwenberg (2022) to multiple players and allowing for heterogeneity in the relevance of observers' opinions. This accommodates for the possibility that some coplayers' opinion is more important than others', or even that one may wish to signal opposite traits to different audiences (as per opposite-signed weights). The extension to multiple players allows us make predictions and to investigate an understudied feature of image concerns: how the emotional valence of reputations varies with the number of observers holding some belief. Within the framework developed, if observers do affect behaviour, then the number of (identical) passive onlookers should be a relevant determinant of the resulting impact of audiences on prosocial behaviour.

Charity Dictator mini-Game with External Audience To answer two nested questions on social influence in economic games I develop and test a Dictator Game with dichotomous action, a charity organization as co-player, and a passive external Audience (more on this in Section 3). The first question is whether audience effects as defined earlier *can induce prosocial behavior* through image concerns. The second one is whether audience effects *depend on the number of passive observers* (the audience size). The main contribution of this work then strands from pairing a clean definition of audience effects and image concerns (reflected in the theoretical model) with a stark experimental design with a minimal, neat manipulation, aimed at minimizing potential confounders and granting internal validity to the identification.

In the laboratory experiment, dictators make a binary choice between whether or not to donate half of their endowment to charitable organizations (which play the role of the Recipient in a standard Dictator Game). This allows us to strip the interaction of any form of ambiguity: donating is a prosocial action, while failing to do so is not. As a consequence, when an audience is present, an action can be interpreted as a clear signal about the Dictators' prosociality. Moreover, as noted in Zizzo (2011), while Dictator Games are generally confounded by "experimenter demand effects" (Zizzo (2010)), they are appropriate to study philanthropic giving and image concerns precisely because they identify agents' response to social pressure (rather than, for instance, their social preferences).

I thus provide novel and reliable evidence on size-dependent image concerns, and propose and test a related model within the existing theoretical framework from Battigalli and Dufwenberg (2022). Providing clean evidence on how a minimal intervention such as *pseudo-observability*[5] from an external, noninteractive audience can induce prosocial behavior (and on this effect holds irrespective of the audience's size) may also help design methodologically cleaner experiments, where this effect is accounted for.

The rest of this thesis is structured as follows: Section 2 presents a general and parsimonious model for size-dependent image concerns, making some simplifying assumptions to bring its hypotheses to the empirical framework; Section 3 describes the experimental design based on a Charity Dictator mini-Game with External Audience; Section 4 presents and discusses the experimental results; finally, Section 5 wraps up the article. The appendix includes additional results and robustness checks, along with the instruction for the participants and other technical details about the experiment.

2 Theoretical Framework

In this section I outline a parsimonious model of Size-dependent Image Concerns that lends itself to natural extensions across various social contexts where actions can be interpreted as signals for imperfectly observable traits. The mathematical framework employed is based on the textbook in Game Theory by Battigalli et al. (2023), while the model follows Section 2.4, "*opinions about good/bad traits*", in Battigalli and Dufwenberg (2022).[6] In models which belong to this class, people care about other's opinion about a "Good" private trait which is heterogeneous in the population. Quite generally, in Battigalli and Dufwenberg (2022) this trait represents agents' intrinsic motivation to "carry out good deeds and avoid bad ones". To influence others' ex-post perception of how intrinsically motivated they are – that is, their reputation – agents are willing to forgo some material payoff. Ceteris paribus, the larger the material payoff they are willing to forgo, the more image-concerned an agent is. This structure is not novel as models on image concerns have evolved in tandem with theories of lying aversion and deception, which share a similar structure (see, among others, Fischbacher and Föllmi-Heusi (2013) and Gneezy (2005)).

[5] Bradley et al. (2018) define pseudo-observability as social environments where information about one's behavior is observed but their actual identity is not.

[6] I also follow Battigalli and Dufwenberg (2022) in simplifying the analysis by limiting the belief-dependency only on beliefs of the first order, which concern behaviours and traits (as opposed to higher order beliefs, which are beliefs about beliefs).

2.1 Concerns about Good traits

Notably, in this framework reputations are valuable per-se even in the absence of material strategic incentives (i.e., future gains from repeated interactions). Because preferences *directly* depend on the beliefs of other, to account for the purely psychological valence of image concerns one must allow utilities to to depend on the terminal, endogenously formed, beliefs of other players. To do so, I adopt the theoretical framework of Psychological Game Theory (hereinafter, PGT) first introduced by Geanakoplos et al. (1989) and later advanced by Battigalli and Dufwenberg (2009) and Battigalli et al. (2019). Because PGT allows to model preferences which directly depend on beliefs, it is the most suitable framework to formally study emotions and other "belief-dependent motivations" (see Battigalli and Dufwenberg (2022) for a vast survey on PGT and its applications).

In their survey, Battigalli and Dufwenberg (2022) introduce the following psychological utility for image concern in the two-players case:

$$u_i(z, \alpha_j; \theta_i^{\mathbf{I}}, \theta_i^{\mathbf{R}}) = \pi_i(z) + \theta_i^{\mathbf{I}} \left[\mathbb{1}_i^G(z) - \mathbb{1}_i^B(z)\right] + \theta_i^{\mathbf{R}} \mathbb{E}\left[\tilde{\theta_i^I} | z; \alpha_j\right] \tag{1}$$

where $\pi_i(z)$ is the material payoff obtained by i in terminal history z. The parameter $\theta_i^{\mathbf{I}} \in \mathbb{R}_+$ represents i's "Good" trait. It quantifies her intrinsic motivation to behave prosocially or, more generally, some other imperfectly observable quality which I will call *virtue*. Then, under the simplifying assumption that all players classify terminal histories z as paths where i's actions are "Good" or "Bad" according to an universally shared standard, $\mathbb{1}_i^G(z)$ is an indicator function equal to 1 if history z belongs to the set of terminal history where i took "Good" actions (respectively, $\mathbb{1}_i^B(z)$ for those where i took "Bad" actions) . In turn, $\left[\mathbb{1}_i^G(z) - \mathbb{1}_i^B(z)\right]$ qualifies the net "Goodness" of i's behaviour in path z.[7]

Assuming perfect but incomplete information (actions are perfectly observed but heterogeneous individual traits and preferences are not), for every terminal history z in the set of all possible terminal histories $\mathcal{Z}$, player j holds an endogenous conditional belief, derived from her conditional probability system, over i's intrinsic motivation trait: $\mathbb{E}\left[\tilde{\theta_i^I} | z; \alpha_j\right]$ – more loosely referred to as her *opinion* about i's trait. In turn, agents have heterogeneous preferences over these terminal beliefs – their social image or reputational concerns. The extent to which player i cares about others' ex-post estimate of $\theta_i^{\mathbf{I}}$ is measured by $\theta_i^{\mathbf{R}} \in \mathbb{R}_+$.

[7]In a sense, intrinsic motivations can also be interpreted as stemming from psychologically valenced self-esteem (see Tesser (1988) and Leary and Baumeister (2000)).

As Battigalli and Dufwenberg (2022) outline, Equation (1) introduces some well-known features of signalling games. Being aware that her actions are a (possibly noisy) signal of her private traits, even if i has weak intrinsic motivations to undertake some costly "Good" actions (low $\theta_i^{\mathbf{I}}$), if she is highly concerned about j's beliefs over her motivations (high $\theta_i^{\mathbf{R}}$), i has incentives to undertake "Good" actions and avoid "Bad". Therefore, there is some material cost that i is willing to incur from altering her behaviour and improve j's opinion of her. Finally note that for this analysis to hold j doesn't need to be an active player, nor a player whose material payoff is in any way affected by the outcome of the game.

This simple and general formulation can then be easily amended to account for many and nontrivial forms of social image concerns, providing sharp and testable predictions. Begin by noting that Equation (1) can be generalized as:

$$u_i(z, \alpha_{-i}; \theta_i^{\mathbf{I}}, \bar{\theta}_i^{\mathbf{R}}) = \pi_i(z) + \upsilon(z; \theta_i^{\mathbf{I}}) + \iota\left(\mathbb{E}\left[\tilde{\theta_i^I}|z; \alpha_{-i}\right]; \bar{\theta}_i^{\mathbf{R}}\right) \tag{2}$$

where $\upsilon(\cdot)$ is a real-valued *intrinsic virtue function* fully characterizing the utility of the (net) "Goodness" of i's actions in every terminal history z. How this "Goodness" is defined can be context-dependent and socially constructed. The flexibility and generality of the model pivots around this. Crucially, $\upsilon(\cdot)$ also defines the type of image concerns agents have: if it captures prosociality, then agents are concerned about how altruistic they appear in the eyes of others (Bénabou and Tirole 2006); if it measures deviations from a social norm, the concern is over how willing to conform to the group they seem (see Bernheim (1994) and Andreoni and Bernheim (2009)); if it represents despicable behaviour people might want to avoid shame (Tadelis (2011)); if it stands for intellectual challenges overcame, they wish to show how smart they are, and so on. [8] The utility structure in 2 thus represents many complex forms of social preferences which share a similar structure, carrying them into the rich framework of PGT.

Image concerns are modelled by the real-valued *social image function* $\iota(\cdot)$, parameterized by $\bar{\theta}_i^{\mathbf{R}}$ – now a vector of parameters $\theta_{ij}^{\mathbf{R}}$ for the ex-post belief of every coplayer j over i's intrinsic motivation. Note how the psychological utility of i's reputation in Equation (2) depends on the ex-post belief over $\theta_i^{\mathbf{I}}$ of multiple coplayers. Quite naturally, if i does not care equally about others' opinion of her, the importance of each such opinion is heterogeneous across coplayers.

[8] Here I take social norms, as well as preferences over virtues or good actions as exogenously given rather than endogenously determined as outcomes of strategic interactions.

More generally, the representation of image concerns in Equation (2) makes explicit the signalling nature of interactions where reputation matters. It does so in a versatile manner that allows for seamless adaptation to diverse forms of social influence, contextual settings, or heterogeneous coplayers that could influence preferences over the trait one wants to signal. By varying $v(\cdot)$ one can consider different *virtues* that agents may wish to signal, including the opposite of another virtue: in some contexts or in the eyes of some audiences people may want appear selfish, antisocial, or nonconformist. One empirical example of this are the results from Moisan et al. (2023), where students observed by future cohorts acted more selfishly as compared to unobserved ones. Similarly, people often choose to remain anonymous when making both extremely meager or generous donations to avoid deviating from some social norm (Raihani 2014). Culture also shapes what constitutes a desirable social image. In Muslim societies, where seeking recognition for one's generosity is stigmatized by Islamic values, making donations publicly visible can decrease donations, as in a field experiment by Lambarraa and Riener (2015). Similarly, college students tend to miss-represent their views on politically sensitive topics by pandering to more socially acceptable positions when they are observed by fellow students (Braghieri 2021). Those and other context-dependent motivations are often accounted for, in the different models, by varying $v(\cdot)$ in the specific signalling game induced by the interaction modelled by the analyst.

2.2 A simple Application: Donating to a Charity

Let us now consider a simple example to highlight the impact of audience size on image concern for appearing altruistic in the eyes of others. This application will also drive the experimental design. Assume that agent i has a choice between two actions, one is unambiguously perceived by everyone as "Good", while the other perceived as "Bad" relative to the other. Then, assume that everyone else perfectly observe i's choice. This description illustrates a simple environment, one instance of which could be being publicly asked to make a donation to a charitable organization.

For simplicity, I assume that the image concerns stemming from the opinion of each observer j about i's intrinsic motivation to donate are additive, and that each opinion is weighted by a factor

$\theta_{ij}^{\mathbf{R}}$, reflecting the importance that i gives to j's opinion.[9] Then, i's preferences are represented by:

$$u_i(z, \alpha_{-i}; \theta_i^{\mathbf{I}}, \bar{\theta}_i^{\mathbf{R}}) = \pi_i(z) + \theta_i^{\mathbf{I}} \left[\mathbb{1}_i^G(z) - \mathbb{1}_i^B(z)\right] + \sum_{\{j \neq i\} \in N} \theta_{ij}^{\mathbf{R}} \mathbb{E}\left[\tilde{\theta_i^I} | z; \alpha_j\right] \tag{3}$$

Where N is the set of all active and inactive players, including i, who perfectly observe i's actions. To further simplify the analysis, assume that agents are divided into low- or high- types according to their intrinsic motivation: $\theta_i^{\mathbf{I}} \in \{\theta_L^{\mathbf{I}}, \theta_H^{\mathbf{I}}\}$ with $\theta_H^{\mathbf{I}} > \theta_L^{\mathbf{I}} \geq 0$. Assuming that agents like to be perceived as being high-type by everyone else, it holds that $\theta_{ij}^{\mathbf{R}} \geq 0$ for all i and j.[10] Now assume that everyone shares same prior belief over i's type, which is common knowledge $-$ for instance, the ex-ante probability that i is selfish equals the share of low-type individuals in the population, which is known by everyone $-$ and that, after i's donation decision, everyone updates their belief consistently with the rules of conditional probability whenever possible. A realistic auxiliary assumption that I make is that, after observing i's decision to donate, the conditional posterior belief of all observers is such that i is strictly more likely to be high-type than he would be if he chose not to. That is, letting $z = G$ denote a terminal history where i donates and $z = B$ one where she fails to do so we have:

$$\forall j \in N \backslash \{i\}, \quad \mathbb{E}[\tilde{\theta_i^I} | z = G; \alpha_j] > \mathbb{E}[\tilde{\theta_i^I} | z = B; \alpha_j] \tag{4}$$

Under those assumptions, by observing Equation (3) it follows that even a low-type i with little intrinsic motivations to donate may be induced to do so if she is highly concerned about the opinion of others. Because she enjoys psychological reward from being perceived as altruistic, there is some monetary payoff she is willing to forego in order to improve her social image. More importantly, under those restrictive but broadly realistic primitive assumptions over preferences, types, and how beliefs are updated, the incentive to donate is weakly increasing in the size of the audience. Every additional observer carries a non negative contribution to i's utility of being perceived as high-type (or altruistic), thus increasing the payoff from donating to the charity.

Testable Hypotheses The analysis above sketches a simplified and reproducible social environment from which the experimental design is developed. By imposing a dichotomous choice of this form, the experiment induces a mini-Game where one action is unambiguously and universally

[9] A reasonable assumption could be that every additional observer, drawn from a pool of identical audience members, has a decreasing marginal impact on i's utility. Because I cannot observe the distribution of traits in the population, empirically I can only test for the monotonicity of u_i in total the number of observers ($|N|$).

[10] In my experiment, all agents are identical from i's perspective by construction and so, for all j, $\theta_{ij}^{\mathbf{R}} = \theta_i^{\mathbf{R}}$. This allows to empirically test the monotonicity of $\iota(\cdot)$ in the number of observers.

'Good' or 'Bad', and hence becomes a reliable signal for unobserved traits. I will then induce a trade-off between a monetary payoff and a better reputation and compare the share of 'Good' actions across different treatments varying the presence and size of an audience.

According to the theoretical framework outlined, this simple and stark design yields the following testable hypotheses:

H1: The share of individuals choosing the "Good" action is larger when an Audience is present.

H2: The share of individuals choosing the "Good" action is monotonically increasing in the number of Audience members.

In the experimental design, the unambiguously "Good" action is represented by choosing to split a monetary endowment of 10 EUR in equal parts with a charitable organization. The unambiguously bad action is instead to keep 9 EUR while donating only 1 EUR to the charity.

3 Empirical Framework

The empirical strategy of this thesis is based on a laboratory experiment involving human subjects playing a Charity Dictator mini-Game with External Audience, a modified Dictator Game devised to study social image concerns in an unambiguous setting. The design involves a baseline with no audience and treatments with either 2 or 7 audience members observing the choice outcome. The design and analysis plan have been publicly pre-registered before running the experiment.[11] The experiment was carried out at Bocconi University and was approved by Bocconi's Ethics Committee Review. The data was collected and handled in compliance with the European General Protection Regulation (GDPR, 2016/679).

Experimental Design The baseline treatment involves an active role which interacts with a charitable organization. As Dictators in a Dictator Game, players assigned to the active role can decide how to distribute a fixed endowment of 10 EUR between themselves and the charity. As in a mini-Game, the choice set is limited to two actions: an equal split of 5 EUR each or selfish choice where the dictator keeps 9 EUR and leaves 1 EUR to the charity.[12] Compared to interactions with human co-players and a larger choice set, this setup implicitly presents a choice between an

[11] The pre-registration can be found on the OSF registry at https://osf.io/yfhdp

[12] Neutral wording was employed throughout the actual experiment and in the instructions handed to the participants.

unambiguously more selfish action and an altruistic one − namely, a social dilemma. This stark and binary contrast allows us to minimize the uncertainty over the altruistic or prosocial valence of actions, reducing the noise around the signal resulting from a choice. This will be especially relevant when actions are observed by an audience that forms beliefs over the active player and is coherent with the ad-hoc simplified theoretical framework introduced in Section 2.2.[13]

Audience Treatments Beyond the baseline treatment, the experimental design includes one manipulation with two intensity levels (small audience or large audience). This manipulation involves the introduction of a non-interactive external audience consisting of randomly selected participants who have no financial stake in the outcome of the interaction. Regardless of what players do, audience members always earn a fixed amount of 7 EUR.

The experiment involves a between-subjects design, meaning that each participant is assigned to only one treatment. In sessions where the audience is present, Dictators' allocation choices, along with an alphanumeric random ID are shared to a group of either 2 (small audience) or 7 (large audience) passive players. Only this information is disclosed to the audience, while any other subject-specific detail, such as their cubicle number, is not shared (this produces what Bradley et al. (2018) define as *pseudo-observability*). Similarly, active players only know the audience's size, while lacking knowledge about any other characteristic of its components. All players are made aware of this information structure.

These design choices also reflect closely the theoretical framework introduced in Section 2, as outlined in its application to charitable donations in Section 2.2. Assuming that an equal split of the endowment between the active player and the charitable cause is universally and unambiguously perceived as "Good" while the unequal split is similarly perceived as "Bad", imaged-concerned individuals might respond to the presence of an audience by acting selflessly (donating more) to improve the audience's perception of their intrinsic motivation to donate. Empirically, this translate into the prediction of a higher share of equal splits in sessions where an audience is present.

Finally, if audience concerns are monotonically increasing in the size of the audience − as in the additive toy-model of size-dependent image concerns outlined in Section 2.2 − the intensity of

[13] Note that, while this may also matter for introspection or self-image (see, among others, Grossman and Van der Weele (2017) and Battigalli and Dufwenberg (2022)), the induced effect should remain constant across treatments.

the treatment (small vs large audience) should be related to the share of equal splits. Given the audience's anonymity, from the standpoint of the active player all audience members are identical in expectation. Therefore, under the assumptions of Section 2.2, a larger audience should increase the psychological valence of one's reputation, rising the incentive to exhibit "Good" behavior by splitting the endowment equally. Consequently, one should expect a higher share of active participants to choose the equal split in treated sessions with an audience size of 7 as compared to sessions with an audience size of 2.

Procedures The experiment took place in April 2023 at the Bocconi Experimental Laboratory for the Social Sciences (BELSS) and involved 270 students from Bocconi University. It was programmed and administered using oTree (Chen et al. 2016) on computer cubicles at BELSS. Participants were recruited through the BELSS's SONA recruitment system among a pool of enrolled students who voluntarily signed up to a "Decision-making experiment". The experiment lasted less than 30 minutes and subjects earned either 5, 7, or 9 EUR for their participation.[14] Upon arrival, participants where directed to their assigned computer cubicles. Once all subjects were seated, instructions where read aloud and then displayed on all monitors (see Section A.2 of the Appendix for the instructions). Following the decision-making phase of the experiment, participants filled-in a brief, non-incentivized questionnaire on basic socio-demographic characteristics (available in Section A.3 of the appendix).

The experiment consisted of a total of ten sessions, six over two consecutive days in April 2023 and four over a single day in October 2023. Each day featured all 3 treatments and the timing of each treatment (morning, noon, or afternoon) was varied across the three days. Each session consisted of 27 participants, BELSS's maximum capacity. The experiment involved a total of 270 subjects, of which 236 were active players. In the baseline treatment there were 81 active players and no audience member. In the small audience treatment there were 75 active players and 6 audience members (2 per session). In the large audience treatment there were a total of 80 active players and 28 audience members (7 per session).

In each session, all participants were assigned to the same treatment group. When the treatment involved an audience, roles were randomly assigned, while in the baseline treatment all subjects

[14] As a positive side effect of the experiment, a total of EUR 786 were donated to the charities involved. See Section A.4 of the Appendix

were active. Participants played 5 rounds under a random stranger matching protocol with the charitable organizations. That is, their allocation choices involved different charities every round, but they were unaware of which charity they were interacting with at any given moment. Participants assigned to the audience role observed all choice outcomes after every round, while active players where not informed about the choices of other participants. At the end of the experiment one round was randomly selected to determine payments, which were carried out in the form of Amazon gift cards. In compliance with BELSS's non-deception requirements, rules and procedures outlined above were carefully explained to all participants (see Section A.2 of the Appendix for the instructions). Participants were informed about the number of participants observing their allocation decision, both in the instructions and through the decision interface, but only in treated Sessions (see Section A.1 of the Appendix for the decision interface).[15]

Finally, to prevent interference from personal preferences over charitable causes or controversies surrounding prominent charitable organizations that could confound the results, a group of five, minor and less-known Italian charitable organizations have been selected as recipients for the donations. These charities operate on various areas including healthcare, elderly care, support for people with disabilities, and youth promotion and engagement. A list of the charitable organizations involved in the study, with a short description for each of them, is available in Section A.2 of the Appendix as part of the instructions handed to the participants.

4 Results

This section presents the results of the experiment. I begin in Section 4.1 by showing some descriptive statistics and making a visual case for the presence of treatment effects which vary with the size of the audience. Yet, owing to some sample unbalances between treatment groups, the main results best emerge in Section 4.2, where the data is analyzed in a random-effects panel regression framework to formally test for the impact of observers on the share of individuals choosing to split their endowment with the charity. The main findings are consistent with the hypotheses derived in Section 2.2. I also assess the presence of gender heterogeneity in the propensity to donate and discuss some robustness checks presented in Section C of the Appendix.

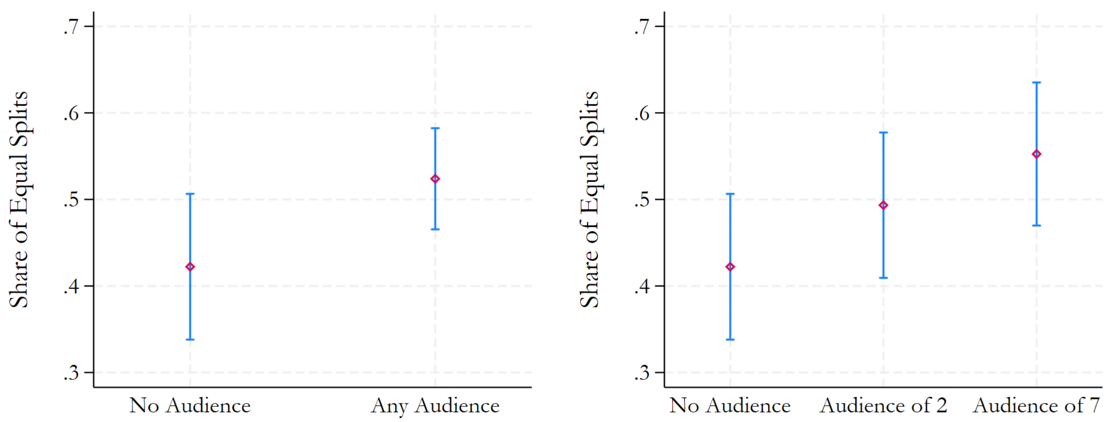
[15]To avoid inducing experimenter-demand effects and nudging participants in the baseline treatment towards selfish behaviour, they were not explicitly told that they were *not* being observed.

4.1 Descriptives and Preliminary Evidence

The data from the experiment suggests that the proportion of equal splits with the charity was higher when subjects were being observed by other participants and that this increase is qualitatively rising in the number of spectators. The average share of equal splits was 42.2% in sessions without an audience, compared to 52.4% when an audience of any size was present (a 10.2 percentage points, or 24.2%, increase from the mean with no audience). Using a t-test on the equality of means between two groups using individual averages (that is, having one observation per individual as their average choice over the 5 rounds), this difference is significant at the 5% confidence level (p-value of 0.0478 on the two-sided hypothesis test, or 0.0239 on the one-sided one in direction predicted by the theory).

Figure 1: Average Share of Equal Splits with 95% Confidence Intervals by Treatment Assignment

(a) Treatment vs Control **(b)** Comparing All Treatments



Note: 95% Confidence Intervals. Both graphs reports the average share of equal splits played by active players. Confidence intervals are computed using individual averages (one observation per participant) across all rounds. Graph (a) on the left pools both groups with a non-zero audience into the "With Audience" category, graph (b) on the right shows each treatment separately.

When comparing the average in each of the two groups with an audience against the average without an audience, only the group with 7 audience members is significantly different from the one without an audience (p-value on the two-sided hypothesis of 0.0295 and of 0.0147 on the one-sided one). For the audience of two, this difference is not statistically significant (p-value of 0.2364 and 0.1182). Yet, this could be a matter of statistical power and, as the treatment is conceptually continuous and the the trend in panel (b) of Figure (1) shows an approximately linear trend, the main specification in Section 4.2 will be a linear regression framework where the number of observers is treated as a continuous variable.

Panel (b) of Figure (1) shows an increasing trend, in line with the hypothesis of size-dependent audience effect. Qualitatively, the frequency of equal splits when an audience of 2 is present is larger than that without any audience, but smaller than the frequency for subjects exposed to audience of 7. This ranking is consistent across all 5 rounds played, as reported in figure B1. This can be interpreted as suggestive evidence that, even for small audiences and mild forms of observability, beyond the presence of an audience, the *number* of observers (the "treatment intensity" in this experiment) might be a significant factor in participants' donation decisions. In Section 4.2 I conduct more rigorous statistical analyses to further explore the validity of this finding.

Table 1: Descriptive statistics and Balance by Treatment. Active Players Only.

Variable	(1) No Audience	(2) Audience of 2	(3) Audience of 7	(4) Difference (2-0)	(5) Difference (7-0)	(6) Difference (7-2)
Age	20.741	20.720	20.663	-0.021	-0.078	-0.058
	(1.571)	(1.849)	(1.534)	(0.274)	(0.245)	(0.272)
Male	0.358	0.507	0.512	0.149*	0.154**	0.006
	(0.482)	(0.503)	(0.503)	(0.079)	(0.078)	(0.081)
# of Experiments	2.136	2.813	2.987	0.678	0.852**	0.174
	(2.042)	(3.502)	(3.087)	(0.455)	(0.412)	(0.529)
Econ or Finance	0.481	0.507	0.312	0.025	-0.169**	-0.194**
	(0.503)	(0.503)	(0.466)	(0.081)	(0.076)	(0.078)
Politics	0.062	0.040	0.075	-0.022	0.013	0.035
	(0.242)	(0.197)	(0.265)	(0.036)	(0.040)	(0.038)
Law	0.074	0.067	0.100	-0.007	0.026	0.033
	(0.264)	(0.251)	(0.302)	(0.041)	(0.045)	(0.045)
Managment	0.358	0.320	0.387	-0.038	0.029	0.068
	(0.482)	(0.470)	(0.490)	(0.076)	(0.077)	(0.077)
Stats or CompSci	0.012	0.027	0.075	0.014	0.063*	0.048
	(0.111)	(0.162)	(0.265)	(0.022)	(0.032)	(0.036)
Other	0.012	0.040	0.050	0.028	0.038	0.010
	(0.111)	(0.197)	(0.219)	(0.025)	(0.027)	(0.034)
Year of Study	2.593	2.573	2.513	-0.019	-0.080	-0.061
	(1.367)	(1.397)	(1.414)	(0.221)	(0.219)	(0.226)
Observations	81	75	80	156	161	155

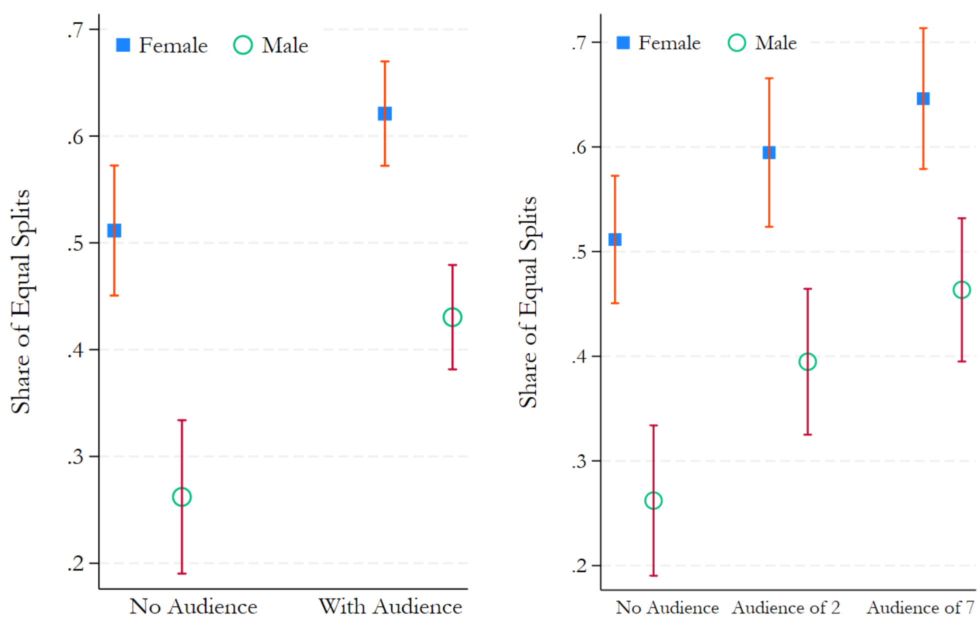
Note: Standard deviations/errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Gender Heterogeneity As will be discussed further in the next section, these means-comparisons across treatment groups conceal significant patterns of heterogeneity across subjects. Most notably, the main determinant of a subject's donation decision in this experiment was their gender: on average, across all treatments men donated a staggering 19 percentage points less than women did, making the latter 1.5 times more likely to donate. This difference is significant at the 1% confidence level using a two sided t-test on individual-level averages (p-value of 0.0001). Because the gender composition of the samples was not perfectly balanced across treatments (see Table 1), those differences are relevant and could confound the preliminary results observed from a simple test of means-comparison. Notably, because there are more males in treated sessions, this unbalance makes it *more* difficult to find evidence in favour of audience effects before conditioning on the participants' gender, favouring the null of no treatment effects. Moreover, significantly more males were assigned to a treatment with some audience present (significant at the 5% confidence level, see Table B.1 of the Appendix). Since males are less likely to donate and over-represented in treated sessions, this unbalance could bias towards 0 estimates derived from a simple comparison between. Thus, these results are best studied conditioning for unbalanced covariates within a regression framework.

Figure (2) outlines the difference in the proportion of equal splits between men and women across treatments. From panel (a) we note that, once conditioning for gender, males experience a larger and more significant increase in the share of equal splits following the introduction of an audience. Panel (b) shows a similar (qualitative) increasing trend for both genders across all treatments: the average share of equal splits rises as the audience gets larger.

Figure 2: Share of Equal Splits with the Charity by Gender and Treatment Assignment

(a) Treatment vs Control **(b)** Comparing All Treatments



Note: 95% Confidence Intervals. Both graphs report the average share of equal splits played by active players. Confidence intervals are computed using individual averages (one observation per participant) across all rounds. Graph (a) on the left pools both groups with a non-zero audience into the "With Audience" category, graph (b) on the right shows each treatment separately.

4.2 Regression Analysis: Linear Audience

The dichotomous choice in the Charity Dictator mini-Game yields a binary outcome variable which, in my preferred specification, is studied in a linear probabilistic panel-data framework through a linear regression model. Consistently with out theoretical framework, the audience size is treated as a numerical variable. Since participants played 5 independent rounds under a random stranger matching protocol with a different charity every round, I employ a random-effects estimator to control for unobserved heterogeneity at the individual level.[16] Consistently with the theoretical framework, here I treat audience as a numerical variable and estimate the following equation using a GLS estimator:

$$\text{Choice}_{i,t} = \beta(\text{Audience Size})_{i,t} + \alpha_{i,t} + \gamma X_{i,t} + \varepsilon_{i,t}$$

Where I adopt heteroskedasticity-consistent standard errors clustered at the individual to allow for serial-correlation between choices across rounds for the same participant, level throughout the analysis, although the results are robust to clustering at these level of the experimental session .[17]

A linear model in the number of observers Given the theoretical framework and the qualitative evidence in the previous section, I treat the size of the audience as a numerical variable and assume it has a constant (and thus weakly monotonic) relationship with the probability of choosing the equal split. The linear fit in Figure (B3) of the Appendix summarizes this relationship in an unconditioned linear framework. Table 2 shows the results using a random-effects GLS estimator fitting a linear model in the size of the audience under different specifications, where 'Audience Size' stands for the number of observers and is the main explanatory variable. Notably, the size of the audience is always significant at the 1% confidence level and remains so once controlling for gender (the unbalanced covariate) in column (2).

Because gender is the main predictor of agents' sharing decisions the experimental data (see Section 4.1) and since it is slightly unbalanced across treatment it is safer to adopt column (2), where the analysis is conditioned on gender, as the main specification. There, the point estimate on the average effect of each individual observer on the share of individuals choosing to split their endowment with the charity is of 2.11 percentage points (or a staggering 4.9% average increase from the baseline of

[16] Reassuringly, the ranking of the average share of equal splits was consistent with the evidence from Figure (1) across all rounds, as shown in Figure (B1) of the Appendix or in the static regression for each round in Figure (3).

[17] As mentioned, the time of the day at which each treatment was administrated changed across sessions of the same treatment to mitigate concerns related to cross-correlation betwen participants to the same session.

Table 2: GLS Random-Effects Estimator

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	Share of Equal Splits				
Audience Size	0.0174**	0.0211***	0.0199**	0.0200**	0.0200**
	(0.00818)	(0.00785)	(0.00788)	(0.00795)	(0.00796)
Male		-0.205***	-0.205***	-0.203***	-0.203***
		(0.0465)	(0.0463)	(0.0479)	(0.0480)
# of Experiments			0.0119	0.00788	0.00788
			(0.00778)	(0.00847)	(0.00849)
Year of Study				0.0303*	0.0303*
				(0.0180)	(0.0181)
Field of Study Controls	No	No	No	Yes	Yes
Round FE	No	No	No	No	Yes
Observations	1,180	1,180	1,180	1,180	1,180
# of individuals	236	236	236	236	236
Avg. share in Control	0.4222	0.4222	0.4222	0.4222	0.4222

Note: Robust standard errors clustered at the individual level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

0.42 in the control with no audience). Adding additional controls and saturating the model leaves the point estimates reassuringly unchanged in size and significance. Similarly, the point estimate on males is always large, negative, and significant at the 1% confidence level. The coefficient on the Male dummy is around 10 times larger than the one on the Audience Size, showing how gender plays a leading role in participants' donation decisions. Even if we saturate the model with the number of previous experiments, the year of study, and the field of study, both the number of observers and gender remain large and significant (see column 4). Finally, there results are robust to including round-fixed effects (column 5), to account for the case where the presence of an audience may induce differential trends over time, yielding to an effect driven by a different dynamic from the one assumed here.

Robustness Section C of the Appendix reports alternative model specifications, showing that the results presented so far are robust to clustering standard errors at the level of the experimental session (See Table C.1), employing a nonlinear (Logit) probability model in a random effects framework (Table C.2), or adopting a (linear) multilevel mixed-effects model with random intercepts at the individual- and experimental session-level (Table C.3). In all these alternative specifications the number of observers displays large and significant estimates, consistent with the main findings from Table 2.

4.3 Comparing Treatments as Categorical Variables

Here I also present and discuss the results of replicating the main analysis when comparing treatment groups treated as categorical variables. While this is not aligned with the theoretical framework and the numerical nature of the treatment, it may better reflect the data structure resulting from an experiment with three treatment assignment groups, although I do not expect our sample size to yield significant results for small treatment effects, as those resulting from the small audience treatment. I then introduce two dummies for each treated group and use GLS to estimate the following linear model with random effects at the individual level model:

$$\text{Choice}_{i,t} = \beta_1(\text{Audience of 2})_{i,t} + \beta_2(\text{Audience of 7})_{i,t} + \alpha_{i,t} + \gamma X_{i,t} + \varepsilon_{i,t}$$

where ε are residuals which I cluster at the individual level to allow for serial correlation over time between observations from the same individual. In columns (1-5) I pool together all groups with some observers under the "Audience" category to explicitly test the first hypothesis on audience effects. Columns (6-10) compare each treatments separately. In both cases, the excluded category is the baseline without an audience.

Table 3: GLS Random-Effects Estimator

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable	Share of Equal Splits									
Audience (any size)	0.102**	0.133***	0.124**	0.131***	0.131***					
	(0.0515)	(0.0490)	(0.0496)	(0.0492)	(0.0493)					
Audience of 2						0.0711	0.102*	0.0945*	0.106*	0.106*
						(0.0595)	(0.0568)	(0.0569)	(0.0565)	(0.0566)
Audience of 7						0.130**	0.163***	0.153***	0.156***	0.156***
						(0.0591)	(0.0565)	(0.0569)	(0.0572)	(0.0573)
Male		-0.210***	-0.209***	-0.209***	-0.209***		-0.210***	-0.209***	-0.209***	-0.209***
		(0.0464)	(0.0462)	(0.0479)	(0.0480)		(0.0463)	(0.0461)	(0.0478)	(0.0478)
# of Experiments			0.0116	0.00739	0.00739			0.0114	0.00717	0.00717
			(0.00769)	(0.00830)	(0.00831)			(0.00773)	(0.00839)	(0.00840)
Year of Study				0.0304*	0.0304*				0.0311*	0.0311*
				(0.0182)	(0.0182)				(0.0181)	(0.0181)
Field of Study Controls	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Round FE	No	No	No	No	Yes	No	No	No	No	Yes
Observations	1,180	1,180	1,180	1,180	1,180	1,180	1,180	1,180	1,180	1,180
# of individuals	236	236	236	236	236	236	236	236	236	236

Note: Robust standard errors clustered at the individual level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

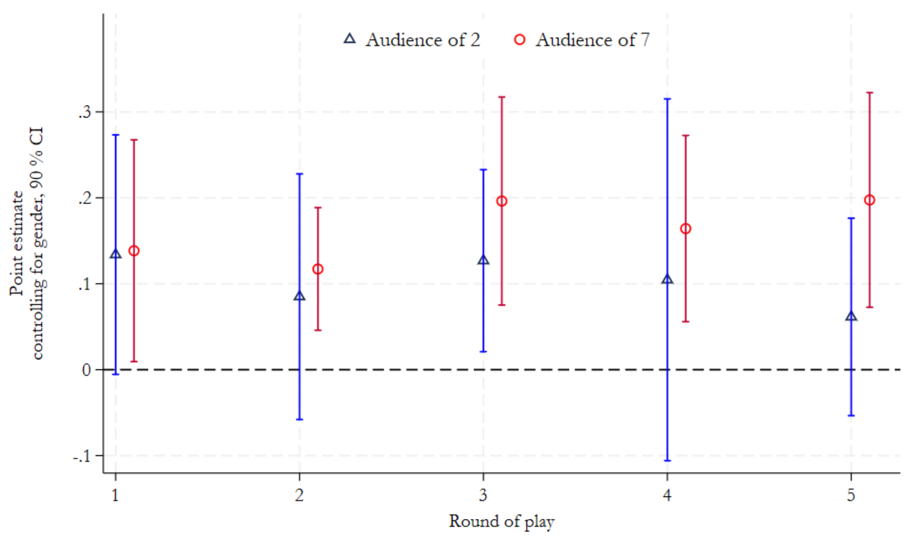
Consistently with what found in the preliminary analysis, the presence of an audience of any size significantly increases the probability of choosing the equal split by an average of 10 to 13 percentage points – an effect comparable in magnitude to that of being a male, which increases donations by around 20 percentage points. Notably, when looking at the estimates for the different treatment groups after conditioning for gender in column (7), the audience of 2 displays a point estimate of 10.2 percentage points which is not statistically different from 0 at the 5% confidence level, while for an audience of 7 this estimate is 16.3 and significant at the 1% confidence level. Although confidence intervals for the two estimates overlap, to the extent to which this it is driven by insufficient power to precisely discern the smaller point estimate in the audience of 2 with the baseline and the two estimates from one another, this differential effect could also be suggestive of size-dependency. Overall, the share of equal splits in the treatment with an audience of 2 is always larger than in the baseline, but this difference is not statistically significant, while the audience of 7 significantly increase donations and displays a larger point estimate.

Round of Play Because participants play multiple round, it could be relevant to assess whether our results are driven by differential trends over time in the different treatment groups. To do so, I use OLS to estimate the following static linear regression model, separately for each round of play:

$$\text{Choice}_i = \beta_1(\text{Audience of }2)_i + \beta_2(\text{Audience of }7)_i + \text{Gender}_i + \varepsilon_i$$

Where $(\text{Audience of }2)_i$ and $(\text{Audience of }7)_i$ are dummy variables for the treatment group. Figure 3 reports the resulting point estimates on the two audience groups for each round of play with 90% confidence intervals. Reassuringly, the point estimate on the audience of 7 is always qualitatively larger than the one for the audience of 2, but their distance seems to grow slightly over time. While this trend is only qualitative, it may hint that audiences of difference sizes may differentially influence whether image concerns *persist*.

Figure 3: Categorical Audience, point estimate from the OLS regression for each round



Note: 90% Confidence Intervals on the Average share of equal splits by treatment assignment and round of play

5 Conclusions

In this work, I studied image concerns in the context of a passive, external audience of anonymous observers. Through a controlled experiment employing a Dictator Game with dichotomous choice and a charity recipient, I provided and discussed novel and reliable evidence on the topic. I also took a conservative approach by introducing a minimal form of observability, referred to as *pseudo-observability* by Bradley et al. (2018), where only actions and random alphanumeric IDs were shared with the audience.

My findings indicate that the presence of an audience can increase the share of equal splits with the charity by 24%, thus promoting prosocial behaviour via image concerns. I also discuss novel evidence suggesting that the size of the audience might have an impact on this effect, highlighting the need for further inquiry.

Finally, this experiment tests and validates predictions derived from a straightforward and simple application of the theoretical framework on Image Concerns proposed by Battigalli and Dufwenberg (2022). By adopting this framework to model size-dependent image concerns and design a clear experiment, I also hope to showcase the value of adopting the theoretical and mathematical framework of Psychological Game Theory in applied research.

References

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond, "Preferences for truth-telling," *Econometrica*, 2019, *87* (4), 1115–1153.

Andersson, Per A., Arvid Erlandsson, Daniel Västfjäll, and Gustav Tinghög, "Prosocial and moral behavior under decision reveal in a public environment," *Journal of Behavioral and Experimental Economics*, 2020, *87*, 101561.

Andreoni, James, "Impure altruism and donations to public goods: A theory of warm-glow giving," *The Economic Journal*, 1990, *100* (401), 464–477.

_ **and B. Douglas Bernheim**, "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects," *Econometrica*, 2009, *77* (5), 1607–1636.

Attanasi, Giuseppe, Pierpaolo Battigalli, Elena Manzoni, and Rosemarie Nagel, "Belief-dependent preferences and reputation: Experimental analysis of a repeated trust game," *Journal of Economic Behavior & Organization*, 2019, *167*, 341–360.

Battigalli, Pierpaolo and Martin Dufwenberg, "Dynamic psychological games," *Journal of Economic Theory*, 2009, *144* (1), 1–35.

_ **and** _ , "Belief-dependent motivations and psychological game theory," *Journal of Economic Literature*, 2022, *60* (3), 833–882.

_ **, Emiliano Catonini, and Nicodemo De Vito**, "Game theory: Analysis of strategic thinking," *Bocconi University, Milan*, 2023.

_ **, Roberto Corrao, and Martin Dufwenberg**, "Incorporating belief-dependent motivation in games," *Journal of Economic Behavior & Organization*, 2019, *167* (C), 185–218.

Bénabou, Roland and Jean Tirole, "Incentives and prosocial behavior," *American Economic Review*, 2006, *96* (5), 1652–1678.

Bernheim, B. Douglas, "A theory of conformity," *Journal of Political Economy*, 1994, *102* (5), 841–877.

Bradley, Alex, Claire Lawrence, and Eamonn Ferguson, "Does observability affect prosociality?," *Proceedings of the Royal Society B: Biological Sciences*, 2018, *285* (1875), 20180116.

Braghieri, Luca, "Political correctness, social image, and information transmission," *Stanford Univiversity Working Papaper*, 2021.

Cañigueral, Roser and Antonia F. de C. Hamilton, "Being watched: Effects of an audience on eye gaze and prosocial behaviour," *Acta psychologica*, 2019, *195*, 50–63.

Charness, Gary and Martin Dufwenberg, "Promises and partnership," *Econometrica*, 2006, *74* (6), 1579–1601.

Chen, Daniel L, Martin Schonger, and Chris Wickens, "oTree—An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, 2016, *9*, 88–97.

Cottrell, Nickolas B, Dennis L. Wack, Gary J. Sekerak, and Robert H. Rittle, "Social facilitation of dominant responses by the presence of an audience and the mere presence of others.," *Journal of Personality and Social Psychology*, 1968, *9* (3), 245.

Cremer, David De and Müriel Barker, "Accountability and cooperation in social dilemmas: The influence of others' reputational concerns," *Current Psychology*, 2003, *22* (2), 155–163.

d'Adda, Giovanna, Martin Dufwenberg, Francesco Passarelli, and Guido Tabellini, "Social norms with private values: Theory and experiments," *Journal of Games and Economic Behavior*, 2020, *124*, 288–304.

Dana, Jason, Roberto A Weber, and Jason Xi Kuang, "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness," *Economic Theory*, 2007, pp. 67–80.

de C Hamilton, Antonia F and Frida Lind, "Audience effects: what can they tell us about social neuroscience, theory of mind and autism?," *Culture and Brain*, 2016, *4* (2), 159–177.

DellaVigna, Stefano, John A List, Ulrike Malmendier, and Gautam Rao, "Voting to tell others," *The Review of Economic Studies*, 2016, *84* (1), 143–181.

Delton, Andrew W, Max M Krasnow, Leda Cosmides, and John Tooby, "Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters," *Proceedings of the National Academy of Sciences*, 2011, *108* (32), 13335–13340.

Dufwenberg, Martin and Martin A. Dufwenberg, "Lies in disguise–A theoretical analysis of cheating," *Journal of Economic Theory*, 2018, *175*, 248–264.

Emler, Nicholas, "A social psychology of reputation," *European Review of Social Psychology*, 1990, *1* (1), 171–193.

Falk, Armin and Andrea Ichino, "Clean evidence on peer effects," *Journal of Labor Economics*, 2006, *24* (1), 39–57.

Fehr, Ernst and Gary Charness, "Social preferences: fundamental characteristics and economic consequences," *CESifo Working*, 2023, (10488).

Filiz-Ozbay, Emel and Erkut Y. Ozbay, "Effect of an audience in public goods provision," *Experimental Economics*, 2014, *17*, 200–214.

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

Geanakoplos, John, David Pearce, and Ennio Stacchetti, "Psychological games and sequential rationality," *Journal of Games and Economic Behavior*, 1989, *1* (1), 60–79.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer, "Social pressure and voter turnout: Evidence from a large-scale field experiment," *American Political Science Review*, 2008, *102* (1), 33–48.

Glazer, Amihai and Kai A Konrad, "A signaling explanation for charity," *American Economic Review*, 1996, *86* (4), 1019–1028.

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

_ , Agne Kajackaite, and Joel Sobel, "Lying aversion and the size of the lie," *American Economic Review*, 2018, *108* (2), 419–453.

Grossman, Zachary and Joël J Van der Weele, "Self-image and willful ignorance in social decisions," *Journal of the European Economic Association*, 2017, *15* (1), 173–217.

Izuma, Keise, Kenji Matsumoto, Colin F. Camerer, and Ralph Adolphs, "Insensitivity to social reputation in autism," *Proceedings of the National Academy of Sciences*, 2011, *108* (42), 17302–17307.

Lambarraa, Fatima and Gerhard Riener, "On the norms of charitable giving in Islam: Two field experiments in Morocco," *Journal of Economic Behavior & Organization*, 2015, *118*, 69–84.

Leary, Mark R and Roy F Baumeister, "The nature and function of self-esteem: Sociometer theory," *Advances in Experimental Social Psychology*, 2000, *32*, 1–62.

Moisan, Frederic, Giuseppe Attanasi, Roberta Dessi, and Donald Robertson, "Public Goods and Future Audiences," *Available at SSRN 4471977*, 2023.

Rabin, Matthew, "Incorporating fairness into game theory and economics," *American economic Review*, 1993, pp. 1281–1302.

Raihani, NJ, "Hidden altruism in a real-world setting," *Biology Letters*, 2014, *10* (1), 20130884.

Regner, Tobias, "What's Behind Image? Toward a Better Understanding of Image-Driven Behavior," *Frontiers in Psychology*, 2021, *12*, 614575.

Schram, Arthur, Gary Charness et al., "Social and moral norms in the laboratory," *UCSB Economics Working Paper Series*, 2013.

Soetevent, Adriaan R, "Anonymity in giving in a natural context—a field experiment in 30 churches," *Journal of Public Economics*, 2005, *89* (11-12), 2301–2323.

Tadelis, Steven, "The power of shame and the rationality of trust," *Haas School of Business Working Paper*, 2011, *3* (2).

Tesser, Abraham, "Toward a self-evaluation maintenance model of social behavior," *Advances in Experimental Social Psychology*, 1988, *21*, 181–227.

Triplett, Norman, "The dynamogenic factors in pacemaking and competition," *The American Journal of Psychology*, 1898, *9* (4), 507–533.

Trivers, Robert L, "The evolution of reciprocal altruism," *The Quarterly Review of Biology*, 1971, *46* (1), 35–57.

Zajonc, Robert B., "Social Facilitation: A solution is suggested for an old unresolved social psychological problem.," *Science*, 1965, *149* (3681), 269–274.

Zizzo, Daniel John, "Experimenter demand effects in economic experiments," *Experimental Economics*, 2010, *13*, 75–98.

_ , "Do dictator games measure altruism?," *Handbook on the Economics of Philanthropy*, 2011.

Appendix to "The Impact of Audience Size on Image Concerns: Evidence from a Charity Dictator Game"

Appendix A presents experimental materials including the computer interface faced by participants, their instructions, and questions from the final questionnaire. Appendix A also reports a final tally of the monetary contributions for each charity involved. Appendix B reports additional figures and tables. Appendix C shows alternative model specifications reported as robustness checks.

A Additional Experimental Materials

Section A.1 shows the computer interface employed in the experiment, Section A.2 reports the translated script of the experimental instructions which includes the list of the charitable organizations involved in the experiment, Section A.3 lists the questions from the final questionnaire.

A.1 Subjects' Computer Interface

The following figures present screenshots of the main user interface. Figures A1 and A2 report the decision interface faced by active players in the decision-making stage when, respectively, there is no audience or an audience of 2 (the one for an audience of 7 is analogous). Figure A3 shows the the choice output observed by the audience after each round.

Figure A1: Decision Interface, No Audience

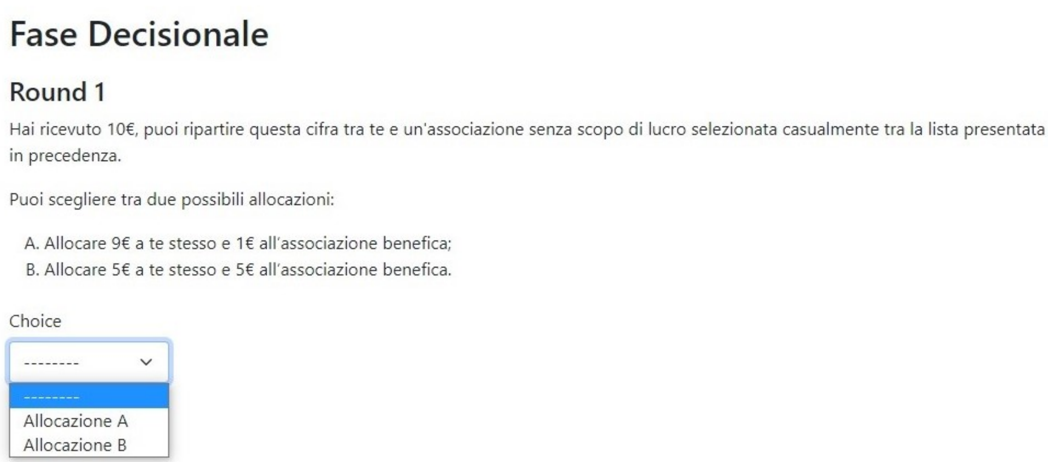


Figure A2: Decision Interface, Audience of 2

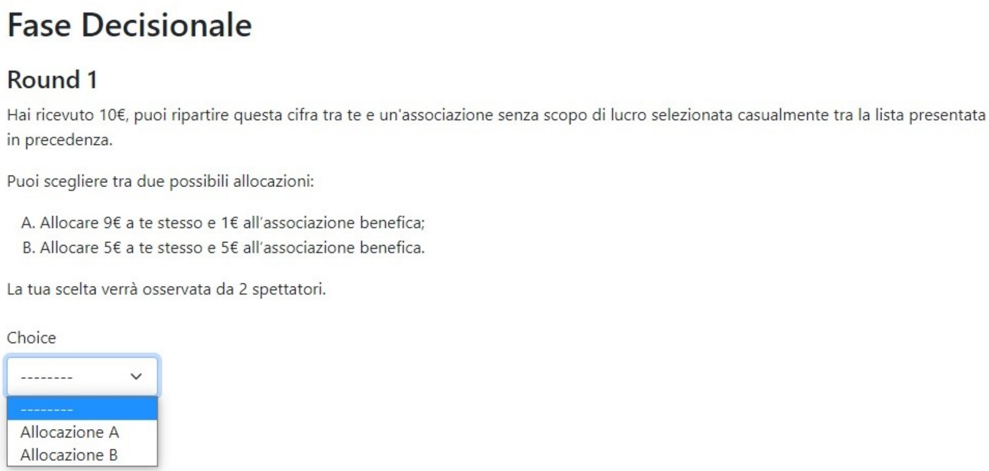
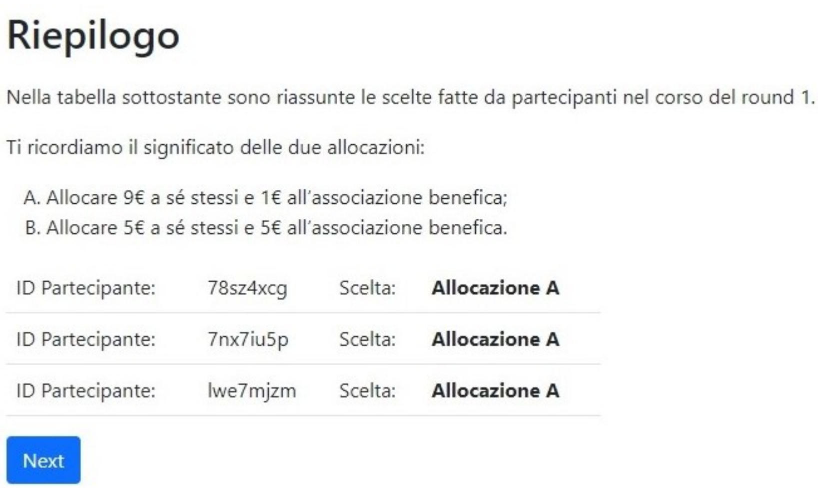


Figure A3: Audience's Feedback



Note: In this example the audience observes the choice output of only three active players.

A.2 Instructions Script

The experiment was conducted in Italian on a pool of Italian-speaking students from Bocconi University. The following instructions have been translated from Italian as literally as possible, so as to preserve the original wording and structure of the sentences. Where an audience was present, only the size of the audience specific to that treatment was displayed. The original instructions in Italian are available upon request.

Instructions (Audience Treatment)

Welcome to BELSS (Bocconi Experimental Laboratory for the Social Sciences). Thank you for taking part in this experiment. Feel free to ask questions by raising your hand. We ask you not to talk to other participants and to keep your cellphones off for the entire duration of the experiment.

The experiment comprises 2 parts: a decision-making stage and a questionnaire. At the beginning of the experiment you will be randomly assigned to an active role or to an audience role. <u>You will keep the same role for the entire duration of the session</u>.

<u>Decision-making stage</u>

The decision-making stage comprises **5 independent rounds**. At the beginning of each round, an active participant will be paired with a different non-profit charitable organization, the list of such organizations is presented below. Each participant will never interact with the same organization twice and the interaction is anonymous; therefore, participants in the active role won't know with which of the 5 organizations they are interacting with at a given time.

At the beginning of each round, every participant in an active role receives and endowment of €10 and decides how to split it between himself/herself and the non-profit charitable organization.

The choice is between two possible allocations:

A. Allocate €9 to yourself and €1 to the charity organization;
B. Allocate 5€ to yourself and 5€ to a charity organization.

The participant in the active role will choose between the two options once per every round, and so 5 times, as many as there are charity organizations that he/she will interact with.

The choice made by the participant in the active role will be **observed by 2[7] participants** in the role of audience members.

<u>All participants are anonymous</u>. The choice made by the participant in the active role will be communicated to the 2[7]observers using a unique code which prevents the identification of the participant by the audience members, both inside and outside the laboratory.

<u>Payment</u>

For every participant in the active role, at the end of the experiment the computer will randomly select one of the 5 rounds. <u>The payment is based on the choice made during the round selected by the computer</u>. This implies that, for every active participant, only the organization involved in the round selected by the computer will receive the amount allocated by the participant.

Participants in the role of audience members will receive a fixed remuneration which does <u>not</u> depend on the choices made by the participant in the active role.

Payments will be carried out by sending Amazon Gift Cards within 15 within working days. To receive the payment is necessary to provide the e-mail address on which you want to receive the gift card.

Instructions (Control Treatment)

Welcome to BELSS (Bocconi Experimental Laboratory for the Social Sciences). Thank you for taking part in this experiment. Feel free to ask questions by raising your hand. We ask you not to talk to other participants and to keep your cellphones off for the entire duration of the experiment.

The experiment comprises 2 parts: a decision-making stage and a questionnaire.

Decision-making stage

The decision-making stage comprises **5 independent rounds**. At the beginning of each round, an active participant will be paired with a different non-profit charitable organization the list of such organizations is presented below. Each participant will never interact with the same organization twice and the interaction is anonymous; therefore, participants in the active role won't know with which of the 5 organizations they are interacting with at a given time.

At the beginning of each round, every participant in an active role receives and endowment of €10 and decides how to split it between himself/herself and the non-profit charitable organization.

The choice is between two possible allocations:

A. Allocate €9 to yourself and €1 to the charity organization;
B. Allocate 5€ to yourself and 5€ to a charity organization.

The participant in the active role will choose between the two options once per every round, and so 5 times, as many as there are charity organizations that he/she will interact with.

Payment

For every participant in the active role, at the end of the experiment the computer will randomly select one of the 5 rounds. The payment is based on the choice made during the round selected by the computer. This implies that, for every active participant, only the organization involved in the round selected by the computer will receive the amount allocated by the participant.

Payments will be carried out by sending Amazon Gift Cards within 15 within working days. To receive the payment is necessary to provide the e-mail address on which you want to receive the gift card.

Questionnaire

At the end of the decision-making stage you will be redirected to a short questionnaire.

List of charity organizations

QuaVio

goals: **to offer care services and support to oncological patients.**

Activities: psychological support; establishment of a relationship of aid, companionship and support; handling administrative paperwork; helping with home transfusions; training caregivers; hygienic care; spiritual assistance; and grief support.

Associazione Arturo Pratelli

Goals: to enhance the **excellence** in education, work and sport for **young people**.

Activities: supporting paths to active and responsible citizenship; organizing fundraisings to support cultural activities and charitable associations.

Le Bollicine

Goals: to promote **sports, rehabilitation and socialization activities for people with disabilities**.

Activities: management and activation of football, fitness and rehabilitative horseback riding courses; supporting the "Casa Clementina" project for autonomous and independent living; organizing summer camps and activities of socialization and integration for people with disabilities.

Auser Comunale

Goals: to promote **active ageing for the elderly and enhance their role in society**.

Activities: help with groceries; Home companionship; social tailoring; transport services for the elderly, disadvantaged or disabled; "Grandparents vigilants" in front of schools; holidays for the elderly; leisure activities; training and refresher training courses; gentle gymnastics classes.

Codini e Occhali ODV

Goals: to **aid the research on rare diseases**, in particular on **Cohen's Syndrome**.

Activities: organizing fundraising and medical-scientific conferences on rare diseases and ADHD; holding an annual day for play, thinking, and awareness involving children and families; cooperating with territorial entities to contribute to the expenses for the care of children; supporting families with ADHD children.

Transparency

All participants to this experiments are entitled to ask the experimenter (by sending an e-mail to alessandro.stringhi@unisi.it) any additional information concerning the non-profit charitable organizations involved in the experiment and the amounts received by them throughout this study.

A.3 Final Questionnaire

After the decision-making portion of the experiment, subjects were asked to fill in an anonymous and non-incentivized short questionnaire on basic socio-demographic characteristics. The original questions in Italian are available upon request.

- Age (in years)
- Gender (male, female, other)
- Field of study (to be selected from a scroll-down menu including: Economics, Finance, or Management; Law; International Relations; Computer Sciences; Mathematics or Statistics; Other)
- Level of Education (year of enrollment)
- Number of experiments in which you took part in the past (integer number)

A.4 Total Contributions to the Charities Involved

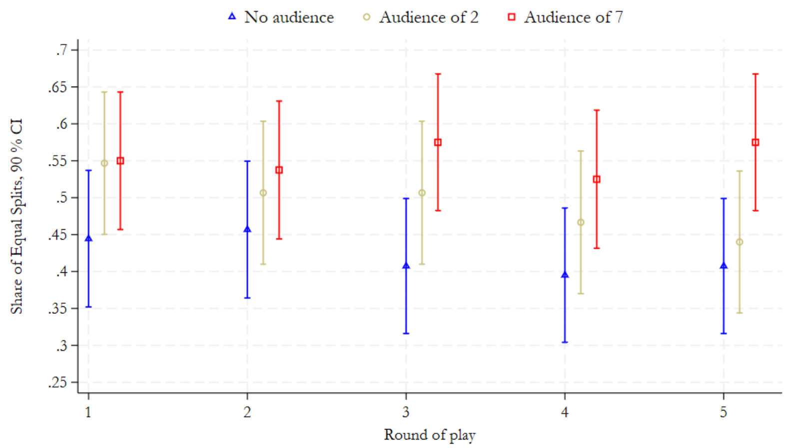
As a positive side-effect of of employing charity co-players in the experiment, a total of EUR 786 were donated to the charity organizations involved. Individual amounts received by each charity vary based on the random selection of both the round selected for payments and the matching between players and charities each every round. The total of € 786 is obtained from the following individual contributions:

- Associazione A. Pratelli: EUR 157
- Auser Comunale (Siena): EUR 144
- Le Bollicine: EUR 180
- Codini e Occhiali ODV: EUR 161
- QuaVio: EUR 144

B Additional Figures and Balance

This Section shows additional and complementary figures and a balance table comparing the control group with both groups with a non-zero audience. Figure (B1) shows the evolution of the proportion of individuals playing the equal share each round, by treatment group. Notably, the proportion of equal splits in the treatment group with an audience of 7 is always above the one for the group with an audience of 2 which is in turn always above that of the control group.

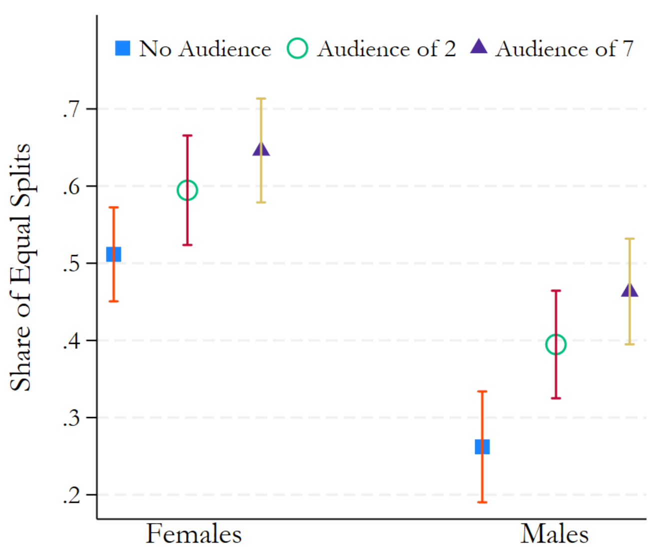
Figure B1: Trendline of the Share of Individuals Choosing the Equal Split by Round and Treatment



Note: 90% Confidence Intervals on the Average share of equal splits by treatment assignment and round of play

Figure (B2) replicates Figure (2) highlighting a comparison of the treatment effect within genders.

Figure B2: CI plot by Gender and Audience Size



Note: 95% Confidence Intervals on the average share of equal splits by Gender and Treatment assignment computed using individual averages (one observation per participant) across all rounds.

Table B.1 shows that sample unbalances are still present when pooling together treatments where an audience is present, highlighting the need to include unbalanced controls in the regression analysis.

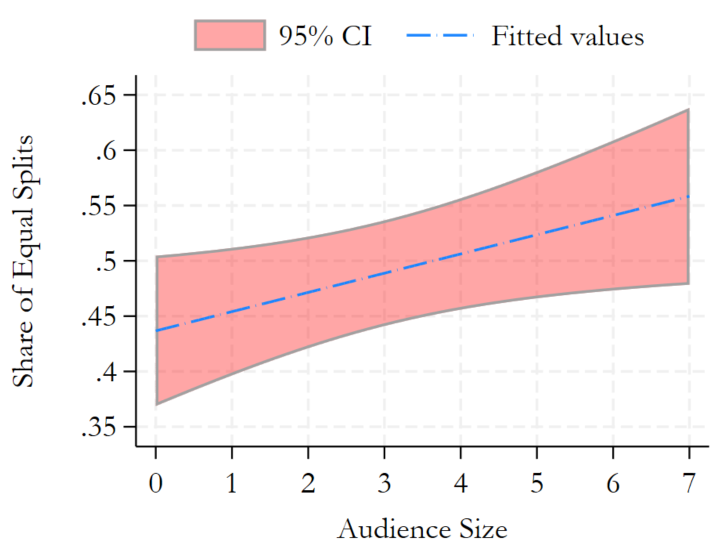
Table B.1: Descriptive statistics and Balance by Treatment. Active Players Only.

Variable	(1) Control (C)	(2) Treatment (T)	(3) Difference (T) - (C)
Age	20.741	20.690	-0.050
	(1.571)	(1.689)	(0.226)
Male	0.358	0.510	0.152**
	(0.482)	(0.502)	(0.068)
Econ or Finance	0.481	0.406	-0.075
	(0.503)	(0.493)	(0.068)
Politics	0.062	0.058	-0.004
	(0.242)	(0.235)	(0.033)
Law Student	0.074	0.084	0.010
	(0.264)	(0.278)	(0.037)
Managment	0.358	0.355	-0.003
	(0.482)	(0.480)	(0.066)
Stats or CompSci	0.012	0.052	0.039
	(0.111)	(0.222)	(0.026)
Other	0.012	0.045	0.033
	(0.111)	(0.208)	(0.025)
Year of Study	2.593	2.542	-0.051
	(1.367)	(1.401)	(0.191)
# of Experiments	2.136	2.903	0.767*
	(2.042)	(3.285)	(0.400)
Observations	81	155	236

Note: Control refers to the treatment group with no audience (the baseline), Treatment refers to the two groups with a non-zero audience (of 2 and 7 respectively). Standard deviations/errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Finally, Figure (B3) shows a linear fit of the share of equal splits in the number of observers.

Figure B3: Linear Fit in the Size of the Audience



Note: 95% Confidence intervals computed using individual averages (one observation per id).

C Robustness – Alternative Model Specifications

This Section reports alternative model specifications, shown as robustness checks. The following tables present estimates produced employing alternative econometric models or specification to analyze the experimental data. The regression output in Table C.1 replicates the main results from Table 2 but clustering standard errors at the level of the experimental session. These estimates are broadly consistent with the ones coming from the baseline and the statistical significance of some estimates increases slightly.

Table C.1: GLS Random-Effects Estimator

	(1)	(2)	(3)	(4)
Dependent Variable	Share of Equal Splits			
Audience Size	0.0174***	0.0211***	0.0199***	0.0200***
	(0.00550)	(0.00559)	(0.00571)	(0.00556)
Male		-0.205***	-0.205***	-0.203***
		(0.0484)	(0.0499)	(0.0485)
# of Experiments			0.0119	0.00788
			(0.00825)	(0.00764)
Year of Study				0.0303
				(0.0202)
Field of Study Controls	No	No	No	Yes
Observations	1,180	1,180	1,180	1,180
# of individuals	236	236	236	236

Note: Robust standard errors clustered by experimental session in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table C.2 shows the results from a Logit nonlinear probabilistic random-effects model, where treating audience size is still treated as a numerical variable standard errors are clustered at the level of the experimental session.

Table C.3 shows that my results still hold, virtually unchanged, also in a multilevel linear regression framework with random intercepts both at the individual level and at that of the experimental session.

Table C.2: Logit Random-Effects Estimator

	(1)	(2)	(3)	(4)
Dependent Variable	Share of Equal Splits			
Audience Size	0.131***	0.159***	0.148***	0.144***
	(0.0482)	(0.0491)	(0.0504)	(0.0488)
Male		-1.553***	-1.561***	-1.577***
		(0.423)	(0.436)	(0.411)
# of Experiments			0.107	0.0715
			(0.0668)	(0.0605)
Year of Study				0.223
				(0.150)
Field of Study Controls	No	No	No	Yes
Observations	1,180	1,180	1,180	1,180
# of individuals	236	236	236	236

Note: Robust standard errors clustered by experimental session in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table C.3: GLS Multilevel Mixed-Effects Estimator

	(1)	(2)	(3)	(4)
Dependent Variable	Share of Equal Splits			
Audience Size	0.0174***	0.0211***	0.0199***	0.0200***
	(0.00550)	(0.00558)	(0.00570)	(0.00554)
Male		-0.205***	-0.205***	-0.203***
		(0.0484)	(0.0498)	(0.0483)
# of Experiments			0.0119	0.00788
			(0.00824)	(0.00761)
Year of Study				0.0303
				(0.0201)
Field of Study Controls	No	No	No	Yes
Observations	1,180	1,180	1,180	1,180
# of individuals	236	236	236	236

Note: Robust standard errors clustered by experimental session in parentheses. *** p<0.01, ** p<0.05, * p<0.1