Prosocial Behavior in Public and Private Spheres: Theory and Experiments*

Fuhai Hong, Pak Hung Lam, Jean Tirole and Chen Zhang

December 8, 2022

Abstract: Technology widens the access to information about our conduct. This paper aims at shedding theoretical and empirical light on induced behavioral changes in our public and private spheres. Agents misallocate efforts between the two spheres by behaving more prosocially in the public sphere than in the private sphere. More importantly, a larger public sphere leads to lower prosociality in both public and private spheres. Overall, giving a socially-valued behavior more visibility does not necessarily make it more prevalent. Two laboratory experiments confirm these findings.

Keywords: Prosocial behavior, multitask signaling, public and private spheres, authenticity, transparency, social score, moral licensing.

JEL numbers: D9, D64, D80, K38.

*Hong: Department of Economics, Lingnan University, Hong Kong; email: fuhaihong@ln.edu.hk. Lam: Division of Social Science, Hong Kong University of Science and Technology; email: phlamae@connect.ust.hk. Tirole: Toulouse School of Economics and Institute for Advanced Study in Toulouse; e-mail: jean.tirole@tse-fr.eu. Zhang: Department of Economics, Lingnan University, Hong Kong; email: chenzhang2@ln.hk.

The authors are grateful to Roland Bénabou, Armin Falk, Karine Van der Straeten, and participants in seminars at the Chinese University of Hong Kong, Princeton University, and Toulouse School of Economics (TSE) for helpful comments, and to Amirreza Ahmadzadeh, Bin Cheng and Paul-Henri Moisson for able research assistance. The Nanjing Audit University and Wuhan University kindly allowed the use of their laboratories. This project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 669217 - ERC MARKLIM). Jean Tirole acknowledges funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, grant ANR-17-EURE-0010. He gratefully acknowledges the financial support of the TSE Digital Center (the list of sponsors is available at https://www.tse-fr.eu/digital). Hong, Lam and Zhang thank the Lingnan University ERSS Fund for financial support. Declaration: No conflict of interest.

1 Introduction

Directly or indirectly, AI, ratings, facial recognition, the recording of online and public-space interactions, and data externalities[1] make our life more and more exposed to public view. The experimental literature's demonstration that we change our behavior when observed by others whose judgment we value suggests that the technological revolution will alter the nature of our social relationships.[2] But if so, how? And how desirable will this transformation be? This paper aims at shedding theoretical and empirical light on these issues.

The inflation of the public sphere implies that behaviors that once belonged to the private sphere are becoming observable to a larger audience.[3] It is easy to predict that the higher visibility of these behaviors and concomitant increase in social pressure will lead us to pay more attention to our reputation. But how will relationships in our remaining private sphere and our overall prosociality be affected?

There is considerable lab-and-field evidence that increased visibility induces more moral behaviors in a wide variety of contexts, from charitable contributions to public goods provision, voting, health and blood donations.[4] This evidence can be summarized in the following assertion: "Giving a socially-valued behavior more visibility makes it more prevalent."

This paper argues that making a behavior more visible may not improve ethics. Consider a multitasking signaling environment in which the individual's trait to be signaled

[1]The sharing by others of information about us on social media, blogs or e-mails.

[2]Vindicating Jaron Lanier (2010) in his assertion that "The most important thing about a technology is how it changes people."

[3]"Public and private spheres", "transparency", and "authenticity" in this paper will have their expected meanings. The *private sphere* will cover actions that are directly observed by a limited set of individuals. This limited observability often characterizes family, friendship and stable work relationships. The *public sphere* by contrast refers to actions observed by a much broader group of onlookers, through public-place behavior, ratings, facial recognition, AI analytics, word-of-mouth, or social networks. How public my behavior is hinges on the number of people who observe my actions and on how much I care about their opinion. *Authenticity* will be defined as the extent to which our behavior reflects our true preferences rather than image concerns. (We focus on social posturing. An individual may not be authentic even if she is not playing a role in relation with others. As has long been acknowledged, people also signal to themselves, as they are watched by their inner spectator (Smith 1759). So, what matters is the degree of authenticity rather than the binary vision of whether individuals are authentic or not). *Transparency*, a reduction in privacy, is the policy through which a given behavior is made more visible to others. A wider audience raises the intensity of our image concerns and affects our behavior.

[4]See e.g. Ashraf-Bandiera (2018) and Bursztyn and Jensen (2017) for overviews of this literature. References include Freeman (1997), Ariely et al. (2009) for charitable contributions, Algan et al. (2016) for public goods provision, Gerber et al. (2008), Funk (2010), DellaVigna et al (2017), Perez-Truglia-Cruces (2017) for voting, Ashraf et al (2014), Karing (2019) for health, and Lacetera et al. (2012) for blood donations. There is also a large experimental literature that manipulates the subjects' self-image concerns and leads to the same conclusion. Finally, in certain environments, what is "socially valuable" depends on peer values and judgment. Experiments on peer pressure (Austen-Smith and Fryer 2005, Bursztyn et al. 2019, Bursztyn et al. 2020a) offer indirect evidence that increased visibility induces more prosocial behaviors, where prosociality is defined relatively to one's in-group and by the (correct or perceived) views of this in-group.

(say, prosociality) is correlated or the same across activities. While increased visibility in one activity bolsters behavior in that activity, the behavioral information educates the audience as to the underlying trait,[5] reducing the scope for signaling in other activities. Thus, making an activity more visible generates both crowding-in and crowding-out in a multi-task environment.

The Theory

Our theory builds on a standard, single-task consensual-behavior model, in which reputational concerns help motivate an agent. In Section 2.1, the latter takes actions that exert externalities onto others. The behavior is consensual in that the audience and the agent agree on what constitutes morally proper behavior. The agent is motivated by (a) her true empathy (or altruism, proclivity for doing good, internalization of others' well-being), (b) her extrinsic motivation, and (c) her image concerns, associated with the inferences that others will draw from their behavior. A behavior is "authentic" if it reflects payoffs (a) and (b), and ignores the social-reputation payoff (c). As is well-known,[6] this benchmark model may exhibit under- or over-provision of prosocial behavior, even when social image is a mere positional good (social prestige is relative and so agents acquire social esteem at the expense of others). Over-provision of the kind envisioned in some dystopian movies and books occurs if and only if the prosocial externality is small;[7] transparency then reduces welfare. For the more relevant case of larger externalities, transparency is socially desirable.

Section 2.2 then analyzes a multi-task extension of the standard model, that exhibits a coexistence between private and public spheres. Behaviors in the private sphere are observed solely through direct interaction, as they cannot be reliably rated or their public disclosure would make the individual or the audience uncomfortable. Behaviors in the public sphere by contrast are the object of public disclosure. There is a two-way interaction between the private and public spheres. First, when behaving prosocially vis-à-vis a public-sphere partner, the agent receives a double dividend: she ingratiates herself with the partner, and she further earns brownie points from third-party onlookers who observe her behavior. This "cheap-signaling effect" is but an extension of the familiar observation that prosociality is encouraged by a widening of the audience. Second, and a novel effect, prosocial behavior in the public sphere signals a minimal level of individual prosociality, and thereby makes it less costly to behave asocially in the private sphere. This is, to the best of our knowledge, the first formalization of the "moral licensing effect" that is

[5]That information about a trait can reduce incentives has been known for a while (e.g. Dewatripont et al 1999). What is novel here is the endogeneity of this trait-relevant information.

[6]Eg. Acquisti et al (2016), Ali-Bénabou (2020), Bénabou-Tirole (2006) and Daughety-Reinganum (2010). At a more abstract level, the possibility that transparency may lead to oversignaling can be traced to the work of Spence on signaling and of Holmström on career concerns.

[7]As illustrated by Lacie in the series *Black Mirror* ("Nosedive", season 3, episode 1). Another instance of over-signaling occurs when people feel compelled to wish "happy birthday" to Facebook "friends" they hardly know (and accept them as "friends" in the first place).

prominent in psychology.[8]

Two main insights emerge. First, prosocial activities, regardless of their overall level, are misallocated, with too much attention paid to the public sphere/too little to the private one. The agent behaves better in the public sphere than in the "all public" or "all private" benchmarks; the converse holds for behavior in the private sphere. Second, the public sphere crowds out the private one. Actually, an expansion in the public sphere (due, say, to technological change) reduces prosociality in both spheres and even reduces overall prosociality over some range. The overall picture is one of public sphere dominance and disintegration of the social fabric in the private sphere.

Overview of Empirical Strategy

In the theory, agents experience one-shot interactions. An agent's reputation's simplest interpretation in this context is pure esteem concerns; the agent longs for self-esteem as well as esteem from others. More generally, the image utility in the model can come from several sources: 1) pure esteem concerns; 2) the prospect of assortative matching or more generally the gains from favorable third-party judgment; and 3) reputation benefit from repeated interactions/reciprocal altruism. Pure esteem concerns are hard to operationalize in a laboratory environment, as individuals remain identified by their number only (self-esteem of course remains but is not directly observable).

We accordingly run two complementary experiments capturing the other two motivations. We robustly test the theoretical predictions by generating reputational benefits from repeated interactions/reciprocal altruism in a dynamic experiment and by creating benefits from a good third-party rating in a static experiment. The dynamic experiment captures motivation 3) by letting subjects interact repeatedly and in a bidirectional manner, while the static experiment only allows subjects to make one-shot, one-way decisions, which are then judged by third parties, thus building on motivation 2). In the dynamic experiment, prosociality is driven by self-image, reciprocal altruism (direct and indirect: I am inclined to be nice to someone who has been nice to me and/or to others), and the reputation concerns arising from other players' reciprocal altruism. In the static experiment, prosociality is driven by self-image as well as the reputation concerns arising from the third parties' indirect reciprocal altruism.[9]

In a basic interaction of either experiment, a dictator decides on whether to take a costly action to help another player, a recipient; the helping behavior exerts a positive

[8]Moral licensing, also called self-licensing, is the phenomenon "whereby increased confidence and security in one's self-image or self-concept tends to make that individual worry less about the consequences of subsequent immoral behavior and, therefore, more likely to make immoral choices and act immorally." (Wikipedia). For example, Monin and Miller (2001) show in their groundbreaking article that when people are made to behave initially in a moral way, they are more likely to display behaviors that are unethical. Effron et al (2009) show that voicing support for Obama in 2008 may license people to make ambiguously racist comments. See also Merritt et al (2012) and Effron et al (2012).

[9]It could be argued that indirect reciprocal altruism is part of motivation 2).

externality on the recipient.

In the dynamic experiment, a population of subjects participated in randomly-matched, pairwise interactions repeatedly.[10] There are 50 rounds in a given treatment and $n = 2k$ subjects, divided into two equal-size subgroups. Each round, each subject participates in the dictator game, either as the dictator or as the recipient. Subjects alternate between being the dictator and the recipient, hence the introduction of two subgroups. In each round, each subject is matched randomly with a subject of the other subgroup, so they have probability $\frac{1}{k}$ of being matched with an arbitrary member of the other subgroup. Matching draws are i.i.d..

The dynamic experiment employs a between-subject design. Subjects are randomly assigned to one of three treatment groups (T0, T40, and T80) indexed by the fraction of interactions that are public/transparent. The baseline treatment, "T0", is an "all private" single-task setting: one's behavior is observable only to those with whom one interacts. In the two multi-tasking treatments, T40 and T80, 40% and 80%, respectively, of the interactions belong to a public sphere, where behavior is recorded in a *social score* that is publicly observable, while the other interactions remain private.[11] The subject's social score is the percentage of contributions in past interactions in the role of the dictator. At the beginning of each round, the dictator learns (a) the interaction history between herself and the recipient, (b) the recipient's social score, and (c) whether the behavior within the round will be used to update the dictator's "social score" (if so, the round is called a public-sphere round).

Note that the only difference between the treatments is that the frequency with which interactions are public. The (random) frequency of future bilateral interactions is left unchanged across treatments; incentives to behave prosocially are thus altered across treatments solely by the presence of the social score—both because the recipient's social score alters the dictator's opinion about the recipient and because the dictator internalizes the consequences of their choice in the round on the updating of their own social score—, not by a modification in the pattern of bilateral interactions. Thus, reciprocal alstruism and the resulting reputation (image) concerns are drivers of prosociality.

In the static experiment, a subject, acting as a dictator, plays a one-shot dictator game towards five recipients through a real-world charity fund. Recipients are passive. The subject's (some) decisions are observed by five third-party observers, who then eval-

[10]The helping game has been used to study the evolution of human cooperation since Nowak and Sigmund (1998).

[11]In 2014, the Chinese government launched a massive policy plan for building a "social credit system" that inter alia would score, publicize and even blacklist fraudulent behaviors or mispractices in marketplace, professions and everyday life. This system's narrative is the development of a unified and numerical record (social score) to evaluate trustworthiness for individuals and businesses. Scoring systems can be found in other countries (e.g., FICO in USA, Schufa in Germany), where most of them were developed by financial institutions to evaluate individuals' creditworthiness. We realize that concerns have been expressed about the social credit system's scoring along divisive issues (see Tirole 2021); here we focus on a social credit system that, as the initial description claimed, aims at inducing better behavior along consensual issues.

uate the subject's generosity; this evaluation determines the subject's income from the experiment. Prosociality is thus driven by, on top of self-image, the subject's reputation concern arising from the observers' indirect reciprocal altruism.

The static experiment mainly employs a within-subject design, in the sense that subjects make decisions in six mutually exclusive worlds, Tx, where $x \in \{0, 1, 2, ..., 5\}$. In world x, there are x recipients and x observers in the public sphere, and $5 - x$ recipients and $5 - x$ observers in the private sphere. Decisions are binary. A prosocial action in the public sphere of world Tx helps x recipients and is observed by all the five observers. A prosocial action in the private sphere of world Tx helps $5 - x$ recipients and is only observed by the $5 - x$ observers in the private sphere. For each action, the helping cost is proportional to the number of recipients being helped. After the subject makes all the ten decisions in the six worlds,[12] one world is randomly drawn as the binding one and the subject's decisions for this world are sent to the observers accordingly for evaluation.

There are pros and cons for each of our dynamic and static experiments. The repeated interactions in the dynamic experiment provide an environment for subjects to learn about the optimal strategy and converge to the equilibrium. The features of repeated, rematched and role-reversing interactions also give the experiment a flavor of realism that is familiar from everyday interactions. Nevertheless, the dynamic nature of the experiment makes it less straightforward to test the theoretical predictions cleanly, which are derived from a static model. In the experiment, prosociality is driven by both the dictator's own reciprocal altruism (when reviewing the recipient's social score and their previous bilateral interactions) as well as the concerns arising from other players' reciprocal altruism. The repeated play of the experiment also inevitably involves learning of social norms. We thus adopt a framework that takes all of these into account and guides us to tease out the confounding factors in testing the theory.

The static experiment, without role-reversal and with the separation between recipients and observers, is simpler. Social learning is absent as subjects make independent one-shot decisions. Recipients do not have any historical records (such as social scores) in the eyes of the dictator and can never take revenge or reciprocate; they are passive. Thus, the dictators' own reciprocal altruism is not invoked. Third-party observers' evaluations do affect the dictator's income. The observers are dis-interested and their evaluations, arguably, are only driven by indirect reciprocal altruism. Subjects' reputation concern arising from the observers' indirect reciprocal altruism is thus the only driver, other than self-image, of the image utility in the experiment. While the one-shot nature of the game and the separation of observers from recipients may not give as much flavor of realism as the dynamic experiment, the static experiment's setting is closer to the theory, which is based on a static model, and enables us to test the theoretical predictions in a relatively straightforward way. We view the two experiments as complementary.

The findings from both experiments support the key theoretical predictions of our model. (1) The cheap-signaling effect generates a higher prosociality in the public sphere

[12]Note that world T0 has no public sphere and world T5 has no private sphere.

than in the all-private treatment. (2) Prosociality in the public sphere decreases with the size of the public sphere. (3) The moral-licensing effect generates a lower prosociality in the private sphere than in the all-private treatment. (4) Prosociality in the private sphere decreases with the size of the public sphere. (5) The subjects misallocate efforts by behaving more prosocially in the public sphere than in the private sphere. (6) Overall prosociality may not increase monotonically with an expansion of the public sphere.

The rest of this paper is organized as follows. The rest of the section reviews the related theoretical and experimental literature. Section 2 develops the theory. Section 3 describes the experimental design and implementation of the dynamic experiment and reports the findings. Section 4 is devoted to the static experiment. The last section concludes.

Related Literature

The model exposited in Section 2 follows the theoretical and empirical paradigm of behavior driven by explicit, implicit and image motivations (Bénabou-Tirole 2006). While this literature emphasizes the role of transparency in incentivizing socially desirable actions, Section 2 qualifies this conventional wisdom.

Through its theme, the paper fits within the broader privacy literature. The case against transparency in the economics literature has several branches. The first branch focuses on abuses by the receiver of the information. Sellers may capture too much of the consumer surplus as they acquire much information about individual tastes (Acquisti et al 2016). They may exploit the consumer's impulsiveness or her incomplete information (people are rarely aware of privacy threats). Information collection may as well destroy insurance (Hirshleifer 1971), most prominently in the realm of health insurance, but also by amplifying the impact of behavioral or information-collection mistakes: subjective profiling ("lazy", "alcoholic"...) may deprive the individual from a job, data dissemination may make a person into a social pariah, etc. Other concerns arising on the receiver side include surveillance by the state and platforms (Tirole 2021) and the violation of the right to be forgotten (the loss of a second chance).

Section 2, which emphasizes the difference in behavior in the public and private spheres, speaks to the multitasking literature (Holmström-Milgrom 1991); in our paper, though, different tasks do not compete for resources (the cost of accomplishing them is additive). Relative to multitask career concerns (Dewatripont et al 1999), the framework puts much more structure and accordingly delivers specific results. Bernheim and Bodoh-Creed (2019) provide a bound on signaling distortions as a function of the number of interactions an agent is engaged in.[13] There are two major differences in focus between

[13]For example, when there is a fixed audience (so relative image concerns tend to zero), the *total* distortion tends to zero under some regularity conditions, reflecting an increase in the signaling efficiency. Bernheim and Bodoh-Creed obtain a general-interest result on the speed of convergence of the signaling distortion, focusing on the separating equilibrium or more generally on equilibria satisfying a dominance

their paper and the analysis of the expansion of the public sphere in Section 2. First, in the latter the public sphere inflates at the expense of the private one and the emphasis is on the impact on behavior in the private sphere. Second, the results hinge on the existence of multiple audiences with different information structures.

Our dynamic experiment is closely related to the experimental literature on helping games that explores the origin and evolution of human cooperation. A standard helping game has the same structure as the social interaction we consider in our theory and experiment. Nowak and Sigmund (1998) theoretically show that cooperation in the form of indirect reciprocity ("give, and you shall be given") could be an evolutionarily stable strategy in a repeatedly played helping game among strangers. The functioning of indirect reciprocity relies on a reputation mechanism that tracks and publicizes a player's behavior in previous interactions.[14] Nowak and Sigmund (1998)'s theoretical result triggered a line of experimental literature that investigates the effects of image scoring in cooperation (e.g. Wedekind and Milinski, 2000; Milinski et al., 2001; Bolton et al., 2005; Seinen and Schram, 2006; Engelmann and Fischbacher, 2009).

The most important difference between our study and this experimental literature lies in our distinction between public and private spheres. The literature focuses on, in our terminology, the single-task, "all public" setting where all behaviors enter social scoring. Our experiment, with co-existence of public and private spheres, allows us to investigate individuals' reallocation of contributive efforts between the two spheres as well as the effect of an expansion in the public sphere.

Our distinction between public and private spheres differs from Engelmann and Fischbacher's (2009) distinction between public and private scores. In their experiment, each subject has a public score in half of the experiment and a private score in the other half. When a subject carries a public score, her entire behavior is observable; when a subject carries a private score, none of her behavior is observable. They use this distinction of public and private scores to separate pure reputation-based indirect reciprocity (i.e. an intrinsic motivation of rewarding good reputation) and strategic reputation building. In contrast, in our dynamic experiment, while every subject has a score that is publicly observable, only behavior in the public sphere is recorded in the score while behavior in the private sphere is not.

There are some other notable differences between our dynamic experiment and the above line of literature. First, in these experiments, typically only a limited history of previous interactions is scored, to mimic humans' limited memory capacity (e.g. Bolton et al., 2005; Seinen and Schram, 2006; Engelmann and Fischbacher, 2009). We are interested in the impact of technological innovation on prosociality. Data technologies make it possible to access unlimited interaction history. Therefore, in our experiment, we let

refinement.

[14] This reputation-based indirect reciprocity is different from another type of indirect reciprocity, called "put-it-foward indirect reciprocity" by Watanabe et al. (2014) or "upstream indirect reciprocity" by Nowak and Sigmund (2005), meaning that a person who has been at the receiving end of a donation may feel motivated to donate in turn (to some third party).

social scores record an accurate summary of the player's previous behavior in the public sphere. Second, since a major motivation of the experimental literature is to test indirect reciprocity, many of these studies (e.g. Wedekind and Milinski, 2000; Milinski et al., 2001; Bolton et al., 2005) focus on interactions among strangers only, in order to shut down direct reciprocity. In our dynamic experiment, we allow for bilateral repeated interactions, which provides signaling incentives in the private sphere. This more realistic setting captures stable relationships with family, close colleagues, friends, etc.

2 Theory

2.1 Single-task benchmark

The model builds on the large theoretical and empirical literature that posits that an individual's social behavior results from her intrinsic motivation to do good for others, her cost of doing so, and finally her desire to project a good image of herself.[15] The benchmark model developed in this subsection is standard.

Drivers of social behavior. There is a continuum of agents with mass 1. Individual i selects an action $a_i \in \{0,1\}$.[16] Action $a_i = 1$ costs the agent $c > 0$ and is pro-social in that it creates an externality $e > 0$ onto the rest of society, while action $a_i = 0$ does not.[17]

Individuals are heterogenous with respect to their desire to do good. Namely, their intrinsic motivation to do good (exert a positive externality) is v, where v is distributed according to smooth cumulative distribution $F(v)$ and density $f(v)$ on $[0, +\infty)$, with mean $\bar{v}$.[18] That the distribution F has support $\mathbb{R}^+$ captures the idea that the behavior is consensual: All agree that $a_i = 1$ is good for the rest of society, although they differ in the extent to which they are willing to incur a cost to contribute. Individual i's intrinsic motivation, v_i, is private information. Individual i cares about others' posterior mean $\hat{v}_i(a_i) = E[v_i|a_i]$ about her type. For the moment, the agent has a single reputation. Later, the audience's information will be heterogenous, and so there will be multiple reputations. This demand for a good reputation may be associated with pure esteem concerns; alternatively, a good reputation allows the individual to derive future benefits from assortive matching or reciprocal altruism. Let μ denote the intensity of image concerns.

[15] The three motivations- intrinsic, extrinsic and image- model is borrowed from Bénabou-Tirole (2006, 2011a) and the broader signaling literature.

[16] Either there is a single action a_i or the agent plays the same action a_i with everyone.

[17] "Externalities" refer to the standard economic notion of inflicting physical harm, raising cost or creating nuisances.

[18] One may argue that a realistic support for F is $[0, e)$, i.e. that agents never put more weight on others than they do on themselves. Note also that assuming that the intrinsic motivation grows with the magnitude of the externality (e.g. can be written ve) would not alter the results. In our experiments, we will take this externality as fixed. We will occasionally state theoretical results for "small" or "high" externalities. These results would however hold under the more general description of intrinsic motivation.

Payoff functions. Agent i's utility is[19]

$$u_i = (v_i - c)a_i + \mu\hat{v}_i.$$

Because u_i is increasing in v_i, there will be a threshold v^* over which the individual behaves prosocially and under which she does not. To define the individual's authenticity, the social pressure is expunged from preferences: Behavior is authentic if $v^* = c$.

Welfare. As long as the reputational payoff is valued for pure esteem concerns, this individual payoff has no social value and reputation is a "positional good": an agent's gain is another agent's loss. And indeed, our assumptions imply that the average reputation is a constant, $\bar{v}$, and so we ignore reputational concerns in W:

$$W = (v_i - c + e)a_i, \tag{1}$$

where e is the social value of the induced externality. Changing the expression of W would affect the regions for over-and under-provision of prosocial behavior (part (ii) of Proposition 1 below), but the qualitative insights would remain the same.[20]

Equilibrium. As already noted, single crossing implies that agent i selects $a_i = 1$ if and only if $v_i \geq v^*$. The cutoff $v^* \equiv v^*(\mu)$, if interior, solves

$$v^* - c + \mu\Delta(v^*) = 0 \tag{2}$$

where[21]

$$\Delta(v^*) \equiv M^+(v^*) - M^-(v^*) \equiv E[v|v \geq v^*] - E[v|v < v^*].$$

We henceforth assume that $1 + \mu\Delta'(v^*(\mu)) > 0$ to preclude any multiplicity of equilibrium. This condition is always satisfied for an "anti-norm" (behaviors are strategic substitutes: $\Delta' > 0$); in the case of a "norm" (behaviors are strategic complements: $\Delta' < 0$), it requires that the intensity μ of image concerns not be too large.

Definition. The *authenticity index* is defined as the ratio $A \equiv \frac{v^*}{c} = 1 - \frac{\mu\Delta(v^*(\mu))}{c} \leq 1$ of the marginal type's intrinsic motivation over the extrinsic one.[22]

[19]Were a_i to be observed by only a fraction x of the population, u_i could be rewritten as $u_i = (v_i - c)a_i + \mu[x\hat{v}_i(a_i) + (1-x)\bar{v}]$. So the same analysis holds, replacing μ by $\tilde{\mu} = \mu x$.

[20]The expression of W could be modified in at least two ways. First, reputation might not be positional (the reputation-stealing game might not be a zero-sum game). Second, the intrinsic motivation v may or may not be part of welfare. The positive results, which will underly the analysis of the experimental results, would not be affected by such variations in the modeling choices.

[21]For a uniform distribution of v on $[0,1]$, $\Delta(v^*) = 1/2$ for all v^*. More generally, Jewitt (2004)'s lemma indicates that (a) if the density f is everywhere increasing, then $\Delta' < 0$; (b) if it is everywhere decreasing, $\Delta' > 0$; and (c) if f is single-peaked, Δ is first decreasing and then increasing in v^*. We will adopt the convention $v^* = 0$ if $-c + \mu\Delta(0) \geq 0$. That is, a corner solution at $v^* = 0$ exists if and only if $\mu\bar{v} \geq c$. Thus, the condition $c > \mu\bar{v}$ is sufficient for the existence of an interior equilibrium.

[22]The authenticity index always lies below 1 for consensual behaviors. This needs not be the case for divisive behaviors (see Tirole 2022).

For example for a uniform density on $[0,1]$, $A = 1 - \frac{\mu}{2c}$. More generally, the stronger the intensity of image concerns (μ), the lower the authenticity, and the more frequent the prosocial behavior.[23]

Comparison with the social optimum. Let us index the socially optimal behavior with superscript "SO". From the expression of W, we see that agent i should choose $a_i = 1$ for all j if $v_i \geq v^{SO}$ (and $a_i = 0$ for all j otherwise), where

$$v^{SO} - c + e = 0.$$

There is underprovision (resp. overprovision) if $v^{SO} < v^*(\mu)$ (resp. $v^{SO} > v^*(\mu)$). Underprovision therefore corresponds to $e > \mu\Delta(v^*(\mu))$.

Proposition 1 *(welfare)*

(i) When faced with an intensity μ of image concerns, individual i picks $a_i = 1$ if $v_i > v^*$ and $a_i = 0$ if $v_i < v^*$, where[24]

$$v^* - c + \mu\Delta(v^*) = 0. \tag{3}$$

(ii) There is underprovision of prosocial behavior if and only if

$$e > \mu\Delta(v^*). \tag{4}$$

(iii) There is more authenticity, the less visible the behavior (the lower μ is).

This proposition checks the standard result according to which there is underprovision of prosocial behavior -too much authenticity- for large externalities and overprovision -too little authenticity- for small ones. The imperfect internalization of the externality is a driver of underprovision, while the desire to gain social recognition may lead to oversignaling for minor externalities. When technology increases image concerns so "prosocial behavior" becomes more frequent, the glory attached to it ($M^+(v^*)$) decreases but truly generous motives pale relative to personal score maximization (the ratio of intrinsic motivation over image concerns decreases).[25]

[23] $\frac{d(\mu\Delta(v^*(\mu)))}{d\mu} = \frac{\Delta}{1+\mu\Delta'} > 0$. This also implies v^* decreases in μ by (2).

[24] See the conditions for interiority in footnote 21.

[25] This is consistent with the second sentence in the following statement of Stuart Russel (2019, page 106): "*[Under a system of intrusive monitoring and coercion] outward harmony masking inner misery is hardly an ideal state. Every act of kindness ceases to be an act of kindness and becomes instead an act of personal score maximization and is perceived as such by the recipient.*" More specific to Section 2.2 is a tentative interpretation of the first sentence, which may be understood as a deterioration of behavior in the private sphere as technology expands the public sphere.

2.2 Multitasking: the dominance of the public sphere

The benchmark model has a single action. The audience may be small or large, but the size of the audience affects only the intensity of the agent's image concerns. A richer picture emerges when the agent chooses multiple actions with different visibility. In particular, some behaviors are bound to remain in the private sphere because they are unobservable by third parties and furthermore cannot be reliably rated.[26] Other behaviors by contrast lend themselves to being shared in the public sphere if the individual, her environment or society decide to disclose them. This section focuses on the mutual interdependence between the private and public spheres, and on how an expansion in the public sphere impacts overall behavior and welfare.

Agent i, with type v_i drawn from distribution $F(v_i)$ with support $\mathbb{R}^+$, interacts with a mass 1 of other agents j [a finite number of interactions (even only two) would not affect the qualitative results]. In each interaction, agent i decides to behave prosocially ($a_{ij} = 1$) or not ($a_{ij} = 0$). Behaving prosocially generates an externality e on agent j (the counterparty) or on society as a whole, and involves private cost c for individual i.

Suppose that a fraction t of individual i's activities is transparent, while a fraction $s = 1 - t$ is private ("t" and "s" stand for "transparent" and "silo"). Individual i knows which activities are transparent or private. In practice, this fraction t may be affected by the technological evolution (cameras, social networks, cheap data storage, artificial intelligence,...), the social pressure for transparency as well as the government's policy or firms' algorithms (see Footnote 11). We will focus on *deterministic symmetric behaviors* within each sphere ($a_{ij} = a_{ik} = a^t$ if j and k are in a public-sphere interaction with i and $a_{ij} = a_{ik} = a^s$ if j and k belong to i's private sphere, and $(a^t, a^s) \in \{0, 1\}^2$). Overall, in a deterministic symmetric equilibrium, agents in i's public sphere observe only a_i^t while agents in i's private sphere observe $\{a_i^t, a_i^s\}$. The single-task model of Section 2.1, which corresponds to $t = 0$ or $t = 1$, is nested in this broader multi-task one.

Assuming that agent i has the same image concern μ with respect to all members in the audience (whether public or private),[27] individual i with type v_i has payoff function

$$u_i = \int_0^1 [(v_i - c)a_{ij} + \mu \hat{v}_{ij}]dj,$$

[26] The reliability of ratings by employers, friends or partners is usually questionable. Outsiders may be unable to ascertain whether a rating within a maintained relationship (or to the contrary following an acrimonious separation) is genuine.

[27] Our analysis can be generalized to different image intensities in the public and private spheres ($\mu^t \neq \mu^s$). Our two key effects, the cheap-signaling effect and the moral-licensing effect, are still present in the generalized model, generating the same misallocation of effort and the same crowding out of prosociality in the private sphere when the public sphere expands. Furthermore, our experiments are designed in such a way that image intensities are identical in the public and private spheres.

where $\hat{v}_{ij}$ is agent i's reputation with agent j. Welfare can be written as

$$W \equiv \int_{i\in[0,1]} \left[\int_{j\in[0,1]} [(v_i - c) + e] a_{ij} dj \right] di$$

Focusing on equilibria that involve deterministic symmetric behaviors with each sphere and letting $\hat{v}_i(a_i^t, a_i^s)$ and $\hat{v}_i(a_i^t)$ denote the posterior expectations of v_i conditional on the information in the private and public spheres, respectively,[28] the authenticity index is equal to $A^t = \frac{v^t}{c}$ in the public sphere and $A^s = \frac{v^s}{c}$ in the private sphere, where v^t and v^s are the cutoffs in the two spheres. Agent i chooses $(a^t, a^s) \in \{0,1\}^2$ so as to solve:

$$\max_{(a^t,a^s)\in\{0,1\}^2} \left\{ (v_i - c)(ta^t + sa^s) + \mu[t\hat{v}_i(a^t) + s\hat{v}_i(a^t, a^s)] \right\}.$$

For expositional conciseness, we rule out corner solutions in the all-private or all-public spheres: For this, we assume that $c > \mu\bar{v}$, so that in the all-private ($t = 0$, $v^s = v^*$, where v^* is the cutoff in the single-task case) or all public ($t = 1$, $v^t = v^*$) cases, not all types contribute ($v^* > 0$). As we will show, there then always exists an equilibrium in which agents behave more prosocially in the public sphere: $v^t < v^s$, as represented in Figure 1. The cutoff in the private sphere is then given by

$$v^s - c + \mu[M^+(v^s) - M(v^t, v^s)] = 0 \tag{5}$$

where $M(v_0, v_1) \equiv [\int_{v_0}^{v_1} v dF(v)]/[F(v_1) - F(v_0)]$ is the expected type given that $v \in [v_0, v_1]$.

Condition (5) captures the moral-licensing effect: Because $M(v^t, v^s) \geq M^-(v^s)$, with strict inequality except when everyone behaves well in the public sphere ($v^t = 0$), condition (5) implies that $v^s \geq v^*$ (where, recall, v^* is the cutoff in the single-task case, given by (3)), with again a strict inequality whenever $v^t > 0$. Even if he does not contribute in the private sphere, he has already separated himself from the chaff if he has contributed in the public sphere.

As for the public sphere, either $v^t = 0$ or $v^t > 0$ is given by the following equation:

[28] We abuse notation by letting $\hat{v}_i$ denote both reputation functions (with one argument when the reputation is in the public sphere and two arguments when it is in the private one). Allowing for deviations from the equilibrium path, and letting $\boldsymbol{a}_i^t$ denote the vector of agent i's actions in the public sphere (that is, $\boldsymbol{a}_i^t = \{a_{ij}\}_{j\in T_i}$ where T_i denote i's public sphere), agent i's objective function can be more generally written as

$$u_i = \int_0^1 \left[(v_i - c) a_{ij} + \mu[\hat{v}_{ij}(\boldsymbol{a}_i^t)\mathbb{1}_{j\in T_i} + \hat{v}_{ij}(\boldsymbol{a}_i^t, a_{ij})\mathbb{1}_{j\notin T_i}] \right] dj$$

letting $\mathbb{1}_{j\in T_i} = 1 - \mathbb{1}_{j\notin T_i}$ denote the indicator function for i's public sphere (equal to 1 if $j \in T_i$ and 0 otherwise).

As usual, one has substantial leeway in specifying off-path beliefs. One can for example take $\hat{v}_{ij} = \hat{v}_i(\min_{k\in T_i} a_{ik})$ in the public sphere and $\hat{v}_i(\min_{k\in T_i} a_{ik}, a_{ij})$ in the private sphere for the functions $\hat{v}_i$ that emerge in the deterministic symmetric equilibrium.

$$t[v^t - c] + \mu\left[t[M^+(v^t) - M^-(v^t)] + s[M(v^t, v^s) - M^-(v^t)]\right] = 0 \tag{6}$$

Condition (6) captures the cheap-signaling effect, implying that $v^t \leq v^*$: The individual perceives an extra reputational payoff, proportional to $\frac{s}{t} = \frac{1-t}{t}$, per good deed in the public sphere. Thus signaling in the public sphere is particularly cheap when the public sphere is small. Besides the standard image benefit $\mu\Delta(v^t)$ per partner in the public sphere, the agent uses the public sphere to engage in damage control in the private sphere.

The analysis of this equilibrium and of its uniqueness is developed in the Appendix. Proposition 2 summarizes the main conclusions.

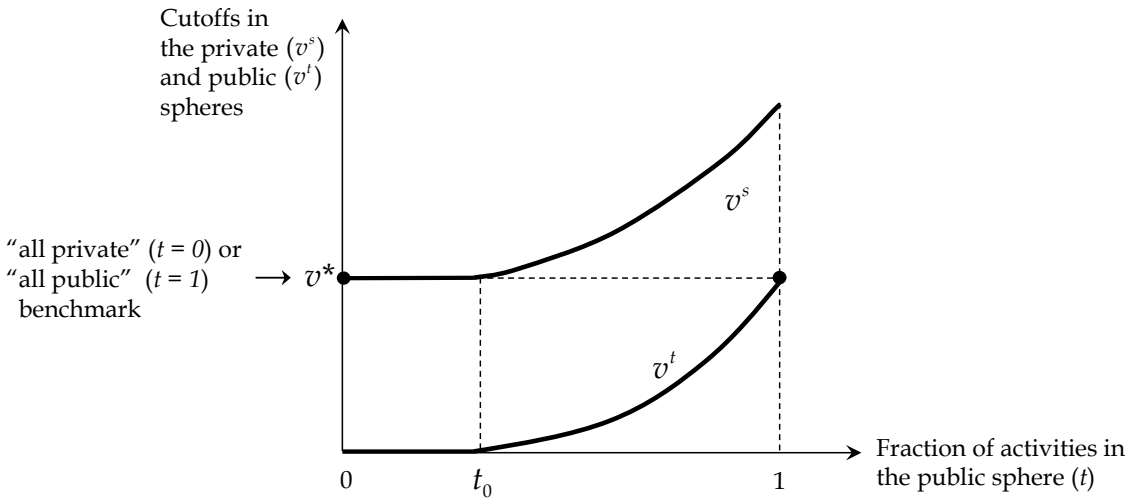


Figure 1: Equilibrium contributions (where $v^* - c + \mu\Delta(v^*) = 0$).

Proposition 2 *(public sphere dominance)*

(i) Existence, uniqueness and monotonicity. There exists an equilibrium satisfying $v^t < v^s$. The cutoffs v^t and v^s are given by conditions (5) and (6), are almost everywhere differentiable in t and satisfy

$$\frac{dv^t}{dt} \geq 0 \quad \textit{and} \quad \frac{dv^s}{dt} \geq 0 \qquad \textit{a.e.}$$

When the density f is non-increasing, this equilibrium is the unique deterministic, symmetric-strategy equilibrium.[29]

(ii) Misallocation. The co-existence of a public and a private spheres implies a misallocation of contributions between the two ($v^t < v^* < v^s$).

[29]When the density f is single-peaked, multiple equilibria may coexist for a small enough public sphere. The monotonicity of v^t and v^s in t however still applies to stable equilibria.

(iii) Crowding out by public sphere. An expansion of the public sphere reduces prosociality in the two spheres. It increases total contribution $\bar{a}(t)$ for $t < t_0$ for some $t_0 > 0$ and over some range reduces the total contribution ($\bar{a}(t)$ decreases with t). $\bar{a}(t)$ is hump-shaped in the case of a uniform distribution of v (and so is welfare if $e > c$).

For a narrow public sphere, the individual acts infrequently in the public sphere and so the visibility/cost ratio is high: Signaling in the public sphere is cheap and $v^t = 0$. Behavior in the public sphere is uninformative and so the individual behaves in the private sphere as in the "all private" benchmark. As t grows, though, signaling in the public sphere becomes more expensive, and this cost effect (weakly) reduces contributions in the public sphere.

The intuition behind Part (ii) of Proposition 2 can be obtained by disconnecting s and t, that is by varying the total number of relationships. First, the existence of a private sphere (making $s > 0$ while keeping $t = 1$ constant) boosts incentives to contribute in the public sphere: If the agent does not contribute in the private sphere, they can still limit the reputational damage in that sphere by contributing in the public sphere. Conversely, the limited reputational damage in the private sphere of not contributing in that sphere, associated with the existence of a public sphere (making $t > 0$ while keeping $s = 1$ constant), crowds out prosocial behavior in the private sphere.

To grasp the intuition behind Part (iii) of Proposition 2, fix the number of relationships ($s + t = 1$). The cost of controlling the damage associated with unethical behavior in the private sphere through ethical behavior in the public sphere increases with the size t of the public sphere and its benefit decreases with the size s of the private sphere. Put differently, damage control is relatively cheap when the public sphere is small, and its cost-benefit ratio increases when behavior becomes more transparent. Thus an expansion of the public sphere discourages contributions in that sphere. Furthermore, having contributed in the public sphere is more of a mark of distinction as the public sphere expands;[30] and so, an expansion in the public sphere crowds out contributions in the private sphere.

The excessive attention to public behavior leads to a disintegration of the social fabric in the private sphere. But the split between private and public sphere also affects the total level of contributions:[31]

$$\bar{a}(t) \equiv s[1 - F(v^s)] + t[1 - F(v^t)],$$

[30] An increase in t reduces contributions in the public sphere. Therefore, an agent who contributes in the public sphere but not in the private sphere has a better image. This reduces the reputation gain of contributing in the private sphere as well (i.e. $[M^+(v^s) - M(v^t, v^s)]$ in (5) is lower).

[31] Welfare is given by $W(t) = t\int_{v^t}^{+\infty}(v - c + e)dF(v) + s\int_{v^s}^{+\infty}(v - c + e)dF(v)$.

And so $\dfrac{dW}{dt} = \left[\int_{v^t}^{v^s}(v - c + e)dF(v)\right] - f(v^t)\dfrac{dv^t}{dt}t(v^t - c + e) - f(v^s)\dfrac{dv^s}{dt}s(v^s - c + e)$.

with $\bar{a}(0) = \bar{a}(1) = 1 - F(v^*)$. Hence, whenever differentiable,

$$\frac{d\bar{a}}{dt} = [F(v^s) - F(v^t)] - sf(v^s)\frac{dv^s}{dt} - tf(v^t)\frac{dv^t}{dt} \tag{7}$$

The first term in the RHS of (7) captures a substitution effect: Contributions are higher in the public sphere and so an expansion of the public sphere raises the overall level of contributions. The other two terms capture the observation that contributions in both spheres decline with an expansion of the public sphere. The overall effect is in general ambiguous. Indeed, for $t \leq t_0$ where $t_0 = \sup(t|v^t = 0)$ (see Figure 1), $\bar{a}(t) = 1 - F(v^*) + tF(v^*)$ is linearly increasing in t; and $\bar{a}(1) = \bar{a}(0) = 1 - F(v^*)$. So $\bar{a}(t)$ must be decreasing over some non-empty range.

The Appendix derives the function $\bar{a}(t)$ in the case of a uniform distribution of v. In that case $\bar{a}(t)$ is hump-shaped: (linearly) increasing on $(0, t_0)$ and decreasing on $(t_0, 1)$.

3 Dynamic Experiment

3.1 Experimental design and hypotheses

A basic interaction in our two experiments is the same as the one in Section 2.[32] In each interaction of the dynamic experiment, the game involves two players: a dictator and a recipient, called player A and player B, respectively, to decontextualize the setting. Player A is endowed with 10 Experimental Currency Units (ECU), and decides whether to contribute the endowment to help player B. If player A contributes the 10 ECU, it will yield 15 ECU to player B. If player A does not contribute, she keeps the 10 ECU and player B receives nothing. Player B does not do anything but observing the choice made by player A. In this setting, $c = 10$ and $e = 15$.

An experimental session consists of an even number $n = 2k$ of subjects (between 26 and 34, so $k \in [13, 17]$), divided into two equal-sized subgroups, and lasts for 50 rounds. At the beginning of each round, each subject is grouped with a subject of the other subgroup by random matching. Each pair of subjects then play the above-mentioned helping game. Subjects play the role of player A and player B in alternate rounds.[33] Since subjects are randomly re-matched in each round, they have a chance to interact with someone else more than once.

Our baseline treatment ("T0") corresponds to the all-private case ($t = 0$) in Section 2.2, which itself depicts the "single-task" case in which reputational payoffs are derived from repeated play within each of the k^2 possible dyads. In an interaction of the baseline

[32] This dictator game is a variant of the standard helping game studied since Nowak-Sigmund (1998) in the experimental literature on indirect reciprocity.

[33] That is, the subjects in one subgroup play the role of dictators in odd rounds and play the role of recipients in even rounds, and the opposite applies for the subjects in the other subgroup.

treatment, when player A makes a decision, she can (only) observe the previous interaction history between herself and the matched player B, if any (including the round number, who made the decision, and the decision made). After player A makes the decision, player B observes the choice as well as her previous interaction history with player A. In this baseline treatment, agent i's actions are observable only to those who directly interact with her (j), and she derives time- and history-dependent reputation utility from them, given the possibility of repeated bilateral interactions. Put differently, there are only bilateral reputations within each possible dyad, as each has no information as to how the other has behaved in their relationships with others.

Besides the baseline treatment, we include two additional treatments similar to multitasking,[34] T40 and T80. In these two treatments, each round of social interactions belongs to either a public sphere or a private sphere. A subject's decision (in role A) made in the public sphere is recorded to update the subject's "social score", which will be made observable to her matched players in the following rounds. Specifically, suppose a subject has been engaged in Y public-sphere rounds as player A (the dictator) and decided to contribute in Z of these rounds; the ratio (i.e., Z out of Y) forms her social score. We do not provide any economic incentive related to social scores other than observability. In the private sphere, however, past decisions made by both parties in the corresponding dyad are only revealed within the dyad, as in the baseline treatment.

At the beginning of each round of interactions in treatments T40 and T80, players are notified of (a) their current match, (b) the interaction history between the two matched players, (c) the matched player's social score, and (d) whether the current interaction is in the public sphere or in the private sphere. In treatment T40, 40% of the social interactions are in the public sphere ($t = 0.4$), with the rest being in the private sphere. In treatment T80, 80% of the social interactions are in the public sphere ($t = 0.8$), and the rest are in the private sphere.[35] In contrast, in the baseline treatment (T0), all interactions are in the private sphere ($t = 0$).

The experiment was programmed in z-Tree (Fischbacher, 2007). Since the experiment was conducted in China, experimental materials were in Chinese. Figure A1 in the Online Appendix A shows a screenshot of player A's decision interface in treatment T40/T80 that is translated into English. A sample of translated experimental instructions for treatment T80 is in Online Appendix C.

Figure 1 and Proposition 2 imply the following hypotheses.

Hypothesis 1 (cheap signaling in the public sphere): Players are more likely to contribute

[34]This is not within-round multitasking, as the dictator has only one action in each round. But it is similar to multitasking in that the agent plays private and public actions repeatedly (although not in the same round), and so we will keep the multitasking terminology of the model.

[35]In these two treatments, all interactions in one round within a treatment are in the same (public or private) sphere. We let the computer randomly select 20 rounds to be public in T40 and 40 rounds to be public in T80. The distribution of public and private spheres across rounds was determined at the beginning of the experiment, but remained unbeknownst to the subjects, who discovered round by round whether the round was public or private.

in the public sphere of treatment T40 or T80 than in treatment T0.

Hypothesis 2 (more expensive signaling in the public sphere): Players are less likely to contribute in the public sphere of treatment T80 than in the public sphere of treatment T40.

Hypothesis 3 (moral licensing in the private sphere): Players are less likely to contribute in the private sphere of treatment T40 or T80 than in treatment T0.

Hypothesis 4 (augmented moral licensing in the private sphere): Players are less likely to contribute in the private sphere of treatment T80 than in the private sphere of treatment T40.

Hypotheses 1 and 3 imply Hypothesis 5:

Hypothesis 5 (misallocation in contributions): In treatments T40 and T80, players are more likely to contribute in the public sphere than in the private sphere.

Proposition 2 also predicts a non-monotonic relationship between the extent of the public sphere and overall contributions. Due to the lack of granularity in the extent of the public sphere in the dynamic experiment, we will not test any hypothesis regarding the overall contributions here. We will test this in the static experiment.

3.2 Experimental procedures

We conducted the dynamic experiment in the economics laboratory at Nanjing Audit University in October and November 2020. We ran 5 experimental sessions for each treatment. Each session included 26-34 student subjects depending on the sign-ups. In total 438 subjects, without any previous experience with our experiment or similar experiments, participated. The subjects' majors cover fields in business and management, science and technology, and others. On average a session lasted around 75 minutes. We paid each subject based on their payoffs in 4 randomly selected rounds (2 rounds as player A and 2 rounds as player B), at an exchange rate of 1 ECU being converted to 1 Chinese yuan, in addition to a show up fee of 20 or 30 yuan.[36] On average a subject received 52.44 Chinese yuan (equivalent to 8.12 USD at the time when the experiment was implemented). At the end of each session, subjects were asked to finish a questionnaire concerning demographic information.

3.3 Experimental Results

We learn from the dynamic experiment that a broad notion of reciprocal altruism – in the form of direct and indirect reciprocity, and learning of social norms – is an important

[36]Subjects in the first three sessions (one session for each treatment) received show up fee of 20 yuan. Upon the request of the laboratory manager, we raised the show up fee to 30 yuan for sessions run afterwards.

driver of prosocial behavior in the laboratory. Raw statistical results that do not account for these channels only support Hypotheses 1, 3 (partially) and 5. A regression analysis controlling for these channels supports all the Hypotheses 1-5.

(a) Summary statistics

Figure 2(a) plots the average contribution frequency by groups across time. The dashed line with circles, the solid lines with triangles, and with squares indicate the average contribution frequency for T0, T40, and T80, respectively. In the multitasking groups, T40 and T80, the two lines fluctuate with hikes and falls. For treatment T40, the hikes typically occurred in the rounds under the (less frequent) public sphere, while for treatment T80, the falls occurred in the rounds under the (less frequent) private sphere.[37]

Figure 2(b) and 2(c) separately present the average contribution frequency, in the public sphere and in the private sphere respectively, for treatments T40 and T80, with that in the baseline group as a benchmark. All the lines exhibit a declining trend. This can be explained by weaker signaling incentives in later rounds in both the public and private spheres.[38]

Figure 3(a) presents a bar chart on the average contribution frequency for T40 and T80 in the private sphere and the public sphere, respectively, with that in the T0 group as a benchmark.

We first compare the social score treatments (T40, T80) with the T0 baseline. The Mann-Whitney test conducted at the session level shows that, compared with the contribution frequency in the baseline group (35%), subjects under treatment T40 were more likely to contribute in the public sphere (p-value<0.01), and less likely to contribute in the private sphere (p-value<0.05), supporting Hypotheses 1 and 3. The comparison between treatment T80 and the baseline group supports Hypothesis 1 but not Hypothesis 3: Subjects under treatment T80 were more likely to contribute in the public sphere than in the baseline group (p-value<0.01), but their contribution frequency in the private sphere is not significantly different from the baseline (p-value>0.9).

We then test the impact of public sphere expansion by comparing T40 and T80. The Mann-Whitney test conducted at the session level shows that the subjects contribute significantly more in T80 than in T40, both in the public sphere (p-value<0.1) and in the private sphere (p-value<0.01), which is inconsistent with Hypotheses 2 and 4. However, the following regression analysis shows that accounting for reciprocal altruism, an important driver of prosociality in the laboratory, overturns this result and vindicates Hypotheses 2 and 4.

Finally, we focus on the social-score treatments. In both groups, subjects were more likely to contribute in the public sphere than in the private sphere: 54% vs. 20% in T40

[37]Note that all interactions in one round within a treatment were in the same (public or private) sphere. See Footnote 35.

[38]The declining contribution frequency in the public sphere resulted in declining social scores (cumulative contribution frequencies in the public sphere), as shown in Figure A2 in the Online Appendix.

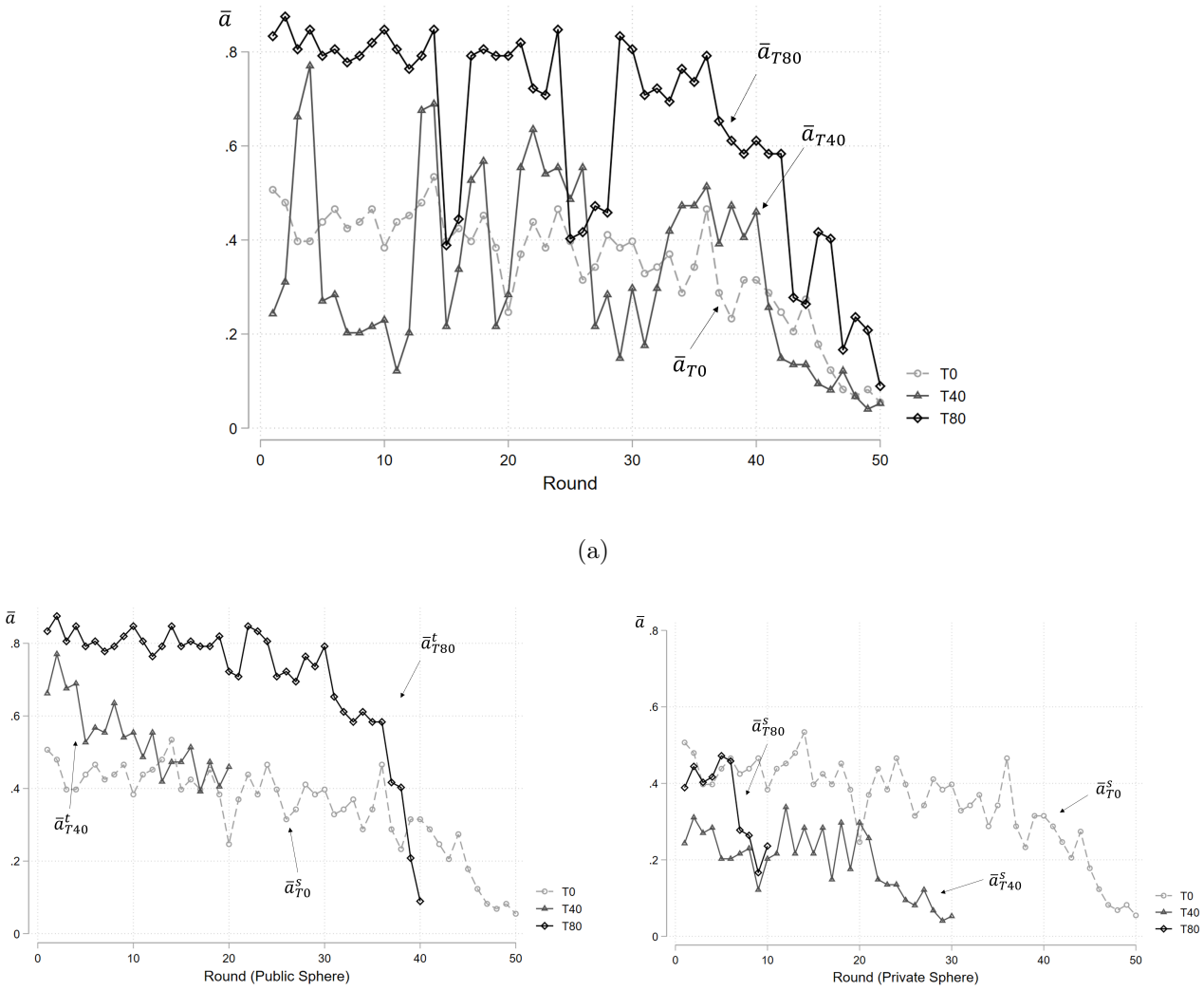
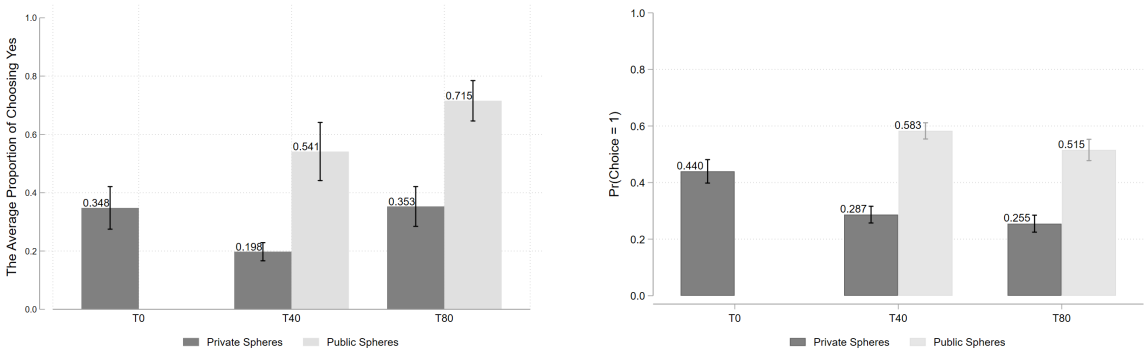


Figure 2: Time trend of contribution frequency

Notes: This figure shows the time trends of the average frequency $\bar{a}$ of making contribution in T0, T40, and T80. In (a), all rounds are included. Figure (b) includes only the rounds in the public sphere for the treatment groups T40 and T80, where the horizontal axis represents the sequence of public rounds; e.g. round 10 means the tenth round in the public sphere. The contribution frequency of group T0 is also presented as a benchmark. Figure (c) includes only the rounds in the private sphere.

Because a given round is either private or public for all subjects, $\bar{a}$ here denotes the contribution in either the private sphere or the public sphere, but not both, while in the theory and empirics elsewhere the notation $\bar{a}$ includes contributions in both spheres.

and 72% vs. 35% in T80 (p-values$<$0.05, Wilcoxon signed rank test at the session level), lending support to Hypothesis 5.



(a) (b)

Figure 3: Average contribution frequency and predicted likelihood of contribution
Notes: Figure (a) shows the average contribution frequency in T0, T40, and T80 in public and private spheres. Figure (b) presents predicted contribution likelihoods under the various scenarios, based on the regression results reported in column (4), Table 1. In calculating the predicted likelihoods, we let the variable of interest (*T40Pub*, *T40Pvt*, *T80Pub*, or *T80Pvt*) be equal to 1 respectively, and specify the values for each of the covariates with their weighted averages adjusted according to their distributions. The error bars represent 95% confidence intervals.

(b) Drivers of behavior

Individuals are inclined to be nice to others who have been nice. This inclination for reciprocal altruism is universal. It also gives rise to reputation concerns in our experimental setting with repeated interactions, because the player who decides whether to contribute in the current round will be judged by others (with reciprocal altruism) in future interactions. Therefore, reciprocal altruism, and the reputation concerns arising from reciprocal altruism, are drivers of prosociality in our experiment. Our theory predicts that the degree of reputation concerns depends on whether the interaction takes place in the public or private sphere and how large the public sphere is. To test these theoretical predictions, our empirical analysis on the drivers of behavior should account for reciprocal altruism.

The dictator's reciprocal altruism depends on her belief on how nice the recepient is. Accouting for this belief updating is arguably complex. One can imagine that there are two types of learning in our setting. One is the learning of the social norm. At the beginning of the experiment, agents may be uncertain about the distribution F of types in the population (the "goodness of society"). They may learn about the social norm (the distribution) from partners' actions and their social scores as the experiment proceeds.[39]

[39]It is worthwhile mentioning that while subjects would naturally learn social norms from others' image scores, the experimental literature on image scoring and indirect reciprocity, reviewed in the Introduction, does not pay much attention to this type of learning. Bénabou-Tirole (2011b) study how the impact of learning on prosociality hinges on the nature of information; in the case of unknown "goodness of society", good news together with strategic complementarities (the definition of a norm) leads to more prosociality. There is also a small empirical literature on the identification of norms and their impact on behavior;

While the theory in Section 2 assumes that the distribution of types is common knowledge, the learning of the social norm when the distribution of types is uncertain may influence contributions.[40] The other type of learning is the learning of the recipient's position in the type distribution, that is about how nice the recipient is relative to the population.

To be specific, given the history observed in previous interactions and the available information on the matched player j, subject i, who acts as the player A, may update the belief on j's type in round τ, $\hat{v}_{ji\tau}$ in the following way:

$$\hat{v}_{ji\tau} = \bar{v}_{i\tau} + \gamma\left(\bar{a}_{ji\tau} - \bar{a}_{i\tau}\right) + \eta\left(s_{j\tau} - \bar{s}_{i\tau}\right) \tag{8}$$

where

$$\bar{v}_{i\tau} = \bar{v} + \kappa\bar{a}_{i\tau} + \lambda\bar{s}_{i\tau}$$

In a nutshell, when estimating j's type, subject i starts from the posterior population mean she learned up to round τ, namely $\bar{v}_{i\tau}$, and uses the relative performance of j both in terms of "j's previous contribution to i" and in terms of j's social score to update her belief. Some explanations are in order. First, the posterior population mean, $\bar{v}_{i\tau}$, consists of three components: (a) $\bar{v}$, a common starting point shared by all subjects when there is no information at the beginning of the experiment; (b) $\bar{a}_{i\tau}$, the frequency of contribution i has received from all the previous matched partners; (c) $\bar{s}_{i\tau}$, the average social score subject i has observed from all her previous matched partners.[41] Second, the relative performance of subject i's current partner j is captured by the two differences, $(\bar{a}_{ji\tau} - \bar{a}_{i\tau})$ and $(s_{j\tau} - \bar{s}_{i\tau})$, where $\bar{a}_{ji\tau}$ is defined as the previous contribution frequency of j, who acted as player A, to subject i, and $s_{j\tau}$ denotes the partner j's social score at round τ. This indicates that subject i updates her belief on j's type by comparing j's observed behavior with her posterior estimate of the population mean.

We now conduct a regression analysis on the subjects' contribution behavior, taking into account reciprocal altruism based on the belief updating framework described in Eq.(8), the decision-makers' demographic characteristics, round fixed effects, etc. Specifically, we employ the following logit model:

$$Prob(a_{ijg\tau} = 1) = \frac{exp(z_{ijg\tau})}{1 + exp(z_{ijg\tau})}$$

where

Besley et al (2015), Chen (2016) and Jia-Persson (2017) find evidence of strategic complementarities in different contexts. In Galbiati et al (2013), subjects learn about the social norm indirectly through the experimenter's choice of incentive rather than from the direct observation of other subjects' behavior. Recently, Bursztyn et al. (2020b) experimentally investigate the learning of social norms and the resulting behavioral changes using Donald Trump's rise in the U.S.

[40] Bénabou and Tirole (2011b, Proposition 3).

[41] The updating of the population mean can be modeled as the following: $\bar{v}_{i\tau} = v_0 + \kappa(\bar{a}_{i\tau} - a_0) + \lambda(\bar{s}_{i\tau} - s_0)$, where v_0, a_0, and s_0 are priors when there is no information. Define $\bar{v} \equiv v_0 - \kappa a_0 - \lambda s_0$; we have $\bar{v}_{i\tau} = \bar{v} + \kappa\bar{a}_{i\tau} + \lambda\bar{s}_{i\tau}$.

$$
\begin{aligned}
z_{ijg\tau} = & [\beta_1 T40Pub_{g\tau} + \beta_2 T40Pvt_{g\tau} + \beta_3 T80Pub_{g\tau} + \beta_4 T80Pvt_{g\tau} + FE_\tau] \\
& + [\beta_5 \bar{a}_{i\tau} + \beta_6 \bar{s}_{i\tau}] \\
& + [\beta_7(\bar{a}_{ji\tau} - \bar{a}_{i\tau}) + \beta_8(s_{j\tau} - \bar{s}_{i\tau})] \\
& + X_i\gamma + e_{ijg\tau}
\end{aligned} \tag{9}
$$

In Eq.(9), $a_{ijg\tau}$ is a binary variable indicating subject i's choice of contribution to matched partner j under treatment group $g \in \{T0, T40, T80\}$ in round τ. With the $T0$ group as the benchmark case, we define dummy variables for treatments and regimes as follows: $T40Pub$ (Public Sphere in treatment T40), $T40Pvt$ (Private Sphere in treatment T40), $T80Pub$ (Public Sphere in treatment T80) and $T80Pvt$ (Private Sphere in treatment T80). FE_τ indicates round fixed effects. Variables $\bar{a}_{i\tau}$, $\bar{s}_{i\tau}$, $(\bar{a}_{ji\tau} - \bar{a}_{i\tau})$ and $(s_{j\tau} - \bar{s}_{i\tau})$ are defined in Eq. (8). Additionally, X_i is a vector of demographic variables of subject i, and $e_{ijg\tau}$ is the error term.[42]

One can interpret Eq. (9) in the following way. The term $[\beta_1 T40Pub_{g\tau} + \beta_2 T40Pvt_{g\tau} + \beta_3 T80Pub_{g\tau} + \beta_4 T80Pvt_{g\tau} + FE_\tau]$ captures "contextual reputational incentives": the incentive to build or preserve one's reputation depends on whether the action will be embodied in the social score, the size of the public sphere, the remaining horizon, and so on. The term $[\beta_5 \bar{a}_{i\tau} + \beta_6 \bar{s}_{i\tau}]$ captures "update of the norm". A norm exists if there is strategic complementarity between others' contributions and subject i's contribution, i.e. if β_5 and β_6 are positive. Finally, the term $[\beta_7(\bar{a}_{ji\tau} - \bar{a}_{i\tau}) + \beta_8(s_{j\tau} - \bar{s}_{i\tau})]$ captures j's relative prosociality standing as perceived in round τ by subject i; this relative standing reflects i's information about both j's over- or under-performance in experienced prosociality in the dyad and j's social score. Put differently, $[\beta_5 \bar{a}_{i\tau} + \beta_6 \bar{s}_{i\tau}]$ in Eq. (9) captures social norm obedience ("I am nice to you as people are nice overall"), and $\beta_7(\bar{a}_{ji\tau} - \bar{a}_{i\tau})$ and $\beta_8(s_{j\tau} - \bar{s}_{i\tau})$ capture direct reciprocity ("I am nice to you as you have been nice to me") and reputation-based indirect reciprocity ("I am nice to you as you have been nice to others") respectively.[43]

Social scores are not well defined in treatment T0. To empirically analyze the data

[42] In our regressions, if we include individual fixed effects, some of the main variables of interest will be dropped, due to multicollinearity. We opt to control for many demographic variables to capture individual characteristics. These demographic variables include age, gender, religiousness, ethnicity (Han Chinese or otherwise), Hukou (household registration, urban or rural), major (business and management, science and technology, or others), year of study, previous experience in any type of economic experiments, annual household income and parents' highest education level. None of the demographical variables shows consistently significant effects.

[43] It is worth mentioning that our notion of direct and indirect reprocity is slightly different from that in the experimental literature on image scoring and indirect reciprocity, which does not consider learning of social norms, as discussed in Footnote 39. The literature typically treats the effect of $\bar{a}_{ji\tau}$ as direct reciprocity and the effect of $s_{j\tau}$ as reputation-based indirect reciprocity (e.g. Nowak and Sigmund, 1998). Watanabe et al. (2014) call the effect of $\bar{a}_{i\tau}$ "put-it-forward indirect reciprocity" ("Someone else has been nice to me and so I am nice to you"), which is narrower than "learning about the norm" in Eq. (9) that also includes learning from $\bar{s}_{i\tau}$.

Table 1: Determinants of contribution behavior: All treatment groups

	(1)	(2)	(3)	(4)	(5)
T40*Pub*	1.975***	3.653***	2.889***	2.430***	2.221***
	(0.488)	(0.500)	(0.463)	(0.272)	(0.281)
T40*Pvt*	0.476***	0.736*	0.534***	0.373***	0.343***
	(0.094)	(0.131)	(0.074)	(0.079)	(0.082)
T80*Pub*	5.120***	1.605**	2.486***	1.595**	1.525*
	(1.284)	(0.300)	(0.443)	(0.332)	(0.350)
T80*Pvt*	1.251	0.321***	0.521***	0.297***	0.284***
	(0.285)	(0.066)	(0.093)	(0.062)	(0.065)
$\bar{a}_{i\tau}$		204.431***	20.816***	29.832***	26.892***
		(79.695)	(4.858)	(6.603)	(5.884)
$\bar{s}_{i\tau}$				2.791**	2.832**
				(1.176)	(1.199)
$(\bar{a}_{ji\tau} - \bar{a}_{i\tau})$		22.598***	19.228***	15.531***	15.439***
		(3.230)	(2.151)	(1.900)	(1.871)
$(s_{j\tau} - \bar{s}_{i\tau})$				9.884***	9.785***
				(2.886)	(2.844)
Wald Test of linear restrictions					
T40*Pub*-T40*Pvt*	4.153***	4.966***	5.410***	6.521***	6.478***
T80*Pub*-T80*Pvt*	4.091***	4.995***	4.773***	5.369***	5.379***
T40*Pub*-T80*Pub*	0.386***	2.276***	1.162	1.524**	1.456**
T40*Pvt*-T80*Pvt*	0.380***	2.289***	1.025	1.255*	1.209
Demographics	Yes	Yes	Yes	Yes	Yes
Round Fixed Effects	Yes	Yes	Yes	Yes	Yes
Pseudo R-squared	0.165	0.400	0.297	0.326	0.323
N	10917	5640	10698	10477	10550
Replacement of Missing Values	No	No	Yes	Yes	Yes

Note: This table reports odds ratios from logit regressions of contribution behavior. Demographic characteristics of the decision maker and round fixed effects are controlled for. Standard errors clustered at the session level are reported in parentheses. *, **, and *** represent significance at 10, 5, and 1% levels, respectively. In column (2), observations with missing values of $\bar{a}_{i\tau}$ or $\bar{a}_{ji\tau}$ are dropped. In columns (3)-(5), we use the values of $\bar{a}_{i\tau}$ to replace the missing values of $\bar{a}_{ji\tau}$. In column (4), we fill in the missing values of $\bar{s}_{i\tau}$ in treatment T0 with the treatment mean of cumulative contribution frequency before the current round; in column (5), the missing values of $\bar{s}_{i\tau}$ in treatment T0 are replaced by the treatment mean of contribution frequency including all rounds. The variable $(s_{j\tau} - \bar{s}_{i\tau})$ takes value 0 in columns (4)-(5) for treatment T0.

from all treatment groups using Eq.(9), one needs to address the missing values of social scores for treatment T0. We take the following two approaches. One is to use certain proxy values to fill in the missing values of social scores in Treatment T0 and include all the three treatment groups in our regressions. The other approach is to include treatment T40 and T80 only, where social scores are present and well-defined, in the regression analysis. This also allows us to account for the differential learning from social scores between treatment T40 and T80, if any. The results are reported in Table 2.

In Table 1, which follows the first approach, column (1) includes only $T40Pub$, $T40Pvt$, $T80Pub$, $T80Pvt$ as the right-hand side variables. Starting from column (2), $\bar{a}_{i\tau}$ and $(\bar{a}_{ji\tau} - \bar{a}_{i\tau})$ are included. Column (4) and (5) further include social scores $\bar{s}_{i\tau}$ and $(s_{j\tau} - \bar{s}_{i\tau})$. The Appendix explains in details how we address the missing values of these four variables in the regressions, while the note of Table 1 concisely reports it.

Table 1 reports odds ratios from the logit regressions of contribution behavior, with standard errors clustered at the session level (following the conventional rule of clustering standard errors at the unit of randomization) and controlling for demographic chracteristics and round fixed effects. The estimates reported in column (1), where only treatment-regime variables are included, are consistent with the pattern shown in Figure 3(a), only partially support Hypothesis 3, and do not support Hypotheses 2 and 4.[44] The odds ratio of $T80Pvt$ is larger than one though not statistically significant. Through the Wald test of linear restrictions, the odds ratios of $T40Pub - T80Pub$ and $T40Pvt - T80Pvt$ are significantly smaller than one. However, after we control for the belief updating from column (2), the regression results support all our hypotheses.

Specifically, starting from column (2), the odds ratios of $T40Pub$ and $T80Pub$ are always greater than one and statistically significant, and the odds ratios of $T40Pvt$ and $T80Pvt$ are always smaller than one and statistically significant. This implies that under the multitasking of public and private spheres, the subjects were more likely to contribute in the public sphere and less likely to contribute in the private sphere, relative to the single-task baseline (T0). Hypotheses 1 and 3 are therefore confirmed. We also find that through the Wald test of linear restrictions, the odds ratios of $T40Pub - T40Pvt$ and $T80Pub - T80Pvt$ are greater than one and statistically significant in all the columns, which confirms Hypothesis 5: under the co-existence of public and private spheres, the subjects were more likely to contribute in the public sphere than in the private sphere.

Moreover, in column (2) and in the most comprehensive specification, column (4), through the Wald test of linear restrictions, the odds ratios of $T40Pub - T80Pub$ and $T40Pvt - T80Pvt$ are greater than one and statistically significant, which implies that with a larger public sphere, the subjects reduced contribution in both public and private spheres, lending support to Hypotheses 2 and 4. In the other comprehensive specification, column (5), the results are qualitatively the same except that the odds ratio of $T40Pvt - T80Pvt$ is greater than one but statistically insignificant (with p-value $= 0.16$). Our theory, which is based on agents' one-shot interactions, assumes that agents derive

[44] It supports the other hypotheses though.

reputation utilities and does not explicitly model the mechanisms by which reputation utilities are possibly generated. In our laboratory setting with repeated interactions, as discussed above, reciprocal altruism is an important driver of prosociality and generates the reputation concern. Therefore, to properly test the theoretical predictions, one has to control for reciprocal altruism in the empirical analysis. After accounting for reciprocal altruism, the regression result overturns the result of the statistical test reported in Section 3.3(a) and supports all the Hypotheses 1-5. In particular, Table 1 shows that adding $\bar{a}_{i\tau}$ and $(\bar{a}_{ji\tau} - \bar{a}_{i\tau})$ to the regression (column (2)-(3)) is crucial in overturning the results. The Appendix further discusses this issue.

The odds ratios of the variables on belief updating are all significantly greater than one. The odds ratios of $\bar{a}_{i\tau}$ and $\bar{s}_{i\tau}$ demonstrate the existence of a norm, i.e. strategic complementarity in contribution: Subjects were more likely to contribute if they were well treated by previous partners or if they learned that the previous partners had contributed often in the public sphere, respectively. The odds ratios of $(\bar{a}_{ji\tau} - \bar{a}_{i\tau})$ and $(s_{j\tau} - \bar{s}_{i\tau})$ show that the subject was more likely to contribute to partners who had been particularly nice to her or in the public sphere, relative to the subject's previous experience or her observation of partners' social scores.

(c) Differential learning of social scores under the public sphere expansion

One might expect the learning from social scores in treatment T80 to differ from the learning in treatment T40, since the public sphere is larger and therefore the social score is more informative under treatment T80. Because there is no learning from $(s_{j\tau} - \bar{s}_{i\tau})$ in treatment T0, it is not possible to differentiate the learning in treatment T40 and that in treatment T80 in a regression with all three treatments. Therefore, we here focus our empirical analysis on the multitasking groups T40 and T80, where social scores are well defined. This approach also avoids arbitrarily assigning "proxy" social scores to treatment T0. We employ the following specification of logit model:

$$Prob(a_{ijg\tau} = 1) = \frac{exp(z_{ijg\tau})}{1 + exp(z_{ijg\tau})}$$

where

$$\begin{aligned} z_{ijg\tau} = {} & \beta_1 Pub_{g\tau} + \beta_2 T80_g + \beta_3 Pub_{g\tau} \times T80_g + \beta_4 \bar{a}_{i\tau} + \beta_5(\bar{a}_{ji\tau} - \bar{a}_{i\tau}) \\ & + \beta_6 \bar{s}_{i\tau} + \beta_7(s_{j\tau} - \bar{s}_{i\tau}) + \beta_8 \bar{s}_{i\tau} \times T80_g + \beta_9(s_{j\tau} - \bar{s}_{i\tau}) \times T80_g \\ & + X_i\gamma + FE_\tau + e_{ijg\tau} \end{aligned} \tag{10}$$

$T80$ is a dummy variable equal to 1 if the interaction occurs in treatment T80, and otherwise 0. Pub is a dummy variable equal to 1 if the interaction occurs in the public sphere, and otherwise 0. Therefore, the baseline case in Eq. (10) is the private sphere under treatment T40.[45] On top of the four variables indicating belief updating for reciprocal

[45]Using $T40Pub$, $T80Pvt$ and $T80Pub$ as defined in Eq. (9) instead of using $T80$, Pub and their

altruism as in Eq. (9), we also introduce interaction terms between $T80$ and two of the variables involving social scores, $\bar{s}_{i\tau}$ and $(s_{j\tau} - \bar{s}_{i\tau})$, to account for the potential differential learning from social scores between treatment T80 and T40.

Table 2: Determinants of contribution behavior: T40 and T80

	(1)	(2)	(3)
Pub	5.562***	6.574***	5.955***
	(1.418)	(1.453)	(1.232)
T80	0.411***	0.746	0.130***
	(0.070)	(0.141)	(0.071)
Pub × T80	1.113	0.919	1.083
	(0.343)	(0.269)	(0.305)
$\bar{a}_{i\tau}$	276.195***	30.782***	31.307***
	(127.704)	(9.652)	(8.765)
$\bar{s}_{i\tau}$	0.585	3.225***	1.406
	(0.344)	(1.373)	(0.560)
$(\bar{a}_{ji\tau} - \bar{a}_{i\tau})$	15.413***	12.853***	12.435***
	(2.136)	(1.077)	(0.927)
$(s_{j\tau} - \bar{s}_{i\tau})$	5.733***	10.508***	5.604***
	(1.397)	(3.382)	(0.789)
$\bar{s}_{i\tau}$× T80			8.814***
			(5.673)
$(s_{j\tau} - \bar{s}_{i\tau})$× T80			6.544***
			(0.879)
Wald Test of linear restrictions			
Pub + Pub × T80	6.189***	6.044***	6.446***
T80 + Pub × T80	0.457***	0.686*	0.141***
Demographics	Yes	Yes	Yes
Round Fixed Effects	Yes	Yes	Yes
Pseudo R-squared	0.423	0.363	0.369
N	3689	6973	6973
Replacement of Missing Values	No	Yes	Yes

Note: This table reports odds ratios from logit regressions of contribution behavior including data from treatment T40 and T80. Demographic characteristics of the decision maker and round fixed effects are controlled for. Standard errors clustered at the session level are reported in parentheses. *, **, and *** represent significance at 10, 5, and 1% levels, respectively. In column (1), we omit the observations if there are missing values. In columns (2) and (3), we use the values of variable $\bar{a}_{i\tau}$ to replace the missing values of variable $\bar{a}_{ji\tau}$.

interaction term in the regression will give us qualitatively the same results.

Table 2 reports odds ratios from logit regressions of contribution behavior with data from treatment T40 and T80, controlling for demographic characteristics and round fixed effects and clustering standard errors at the session level. Column (1) includes $T80$, *Pub*, their interaction term, and the four variables indicating belief updating as in Eq. (9) as right-hand-side variables. We dropped the observations if there are missing values for any of the four belief-updating variables. In columns (2) and (3), we use $\bar{a}_{i\tau}$ to replace the missing values of $\bar{a}_{ji\tau}$ to account for the cases when the dyad was matched for the first time (and so the value of $(\bar{a}_{ji\tau} - \bar{a}_{i\tau})$ is 0 for these cases). Column (3) further introduces the interaction term between $T80$ and $\bar{s}_{i\tau}$ and between $T80$ and $(s_{j\tau} - \bar{s}_{i\tau})$, and so it employs the comprehensive specification as in Eq. (10).

Consistent to the findings from Table 1, the odds ratios of $\bar{a}_{i\tau}$, $\bar{s}_{i\tau}$, $(\bar{a}_{ji\tau} - \bar{a}_{i\tau})$, and $(s_{j\tau} - \bar{s}_{i\tau})$ in columns (1) and (2) are all greater than one and statistically significant except that of $\bar{s}_{i\tau}$ in column (1). After we further introduce the two interaction terms in column (3), they remain greater than one and statistically significant (except the odds ratio of $\bar{s}_{i\tau}$, which is greater than one but statistically insignificant). Interestingly, the odds ratios of the two interaction terms, $\bar{s}_{i\tau} \times T80$ and $(s_{j\tau} - \bar{s}_{i\tau}) \times T80$, are also greater than one and statistically significant, suggesting that learning from social scores in treatment T80 is more effective than in treatment T40.

The regression results are in line with Table 1 and support Hypotheses 2, 4 and 5.[46] In particular, the odds ratios of *T80* and *T80 + Pub × T80* (by Wald test of linear restrictions) represent the effects of the public sphere expansion (from 40% to 80%) on contribution behavior in the private and public spheres, respectively, and they are statistically significantly smaller than one in our most comprehensive specification (column (3)) and in the other columns.[47] These results show that in the presence of a larger public sphere, the subjects reduced contribution in both the private sphere and the public sphere, lending support to Hypotheses 2 and 4. Meanwhile, in all the columns of Table 2, the odds ratios of *Pub* and *Pub + Pub × T80* (by Wald test of linear restrictions) are greater than one and statistically significant, implying that under both treatments T40 and T80, the subjects contributed more in the public sphere than in the private sphere. These results confirm previous findings and support Hypothesis 5. The findings from Table 2 show the robustness of our results after differentiating the learning of social scores in treatments T40 and T80.[48]

Figure 3(b) shows the predicted likelihoods of contribution based on the regression reported in column (4) of Table 1, one of the most comprehensive specification with all the treatment groups. In calculating the predicted likelihoods, we let the variable of interest (*T40Pub*, *T40Pvt*, *T80Pub*, or *T80Pvt*) be equal to 1 respectively, and specify

[46] Table 2 does not include treatment T0 and so cannot test Hypothesis 1 or 3.

[47] One exception is that the odds ratio of $T80$ in column (2) is smaller than one but statistically insignificant (with p-value = 0.12).

[48] The results are qualitatively the same if we further include interaction terms $\bar{a}_{i\tau} \times T80$ and $(\bar{a}_{ji\tau} - \bar{a}_{i\tau}) \times T80$ to column (3) of Table 2; our hypotheses are still supported. Meanwhile, the odds ratios of $\bar{a}_{i\tau} \times T80$ and $(\bar{a}_{ji\tau} - \bar{a}_{i\tau}) \times T80$ are statistically insignificant.

the values for each of the covariates with their weighted averages adjusted according to their distributions (Williams, 2012). Table B1 and Table B2 in the Online Appendix B report the regression results with standard errors clustered at the individual level, which are qualitatively the same as the results reported in Table 1 and Table 2.

4 Static Experiment

4.1 Experimental design and hypotheses

The basic interaction in the static experiment is similar to the one in the dynamic experiment: A dictator decides on whether to contribute an endowment, 20 Chinese yuan, to a recipient. If the dictator contributes, it yields 30 yuan to the recipient. If the dictator does not contribute, she keeps the endowment and the recipient receives nothing. There are five potential recipients, whose roles are passive, and five observers, who will observe some of the dictator's decisions.

We employ both a within-subject design and a between-subject design for the static experiment. For the within-subject design, there are six mutually exclusive worlds, denoted by Tx, where $x \in \{0, 1, 2, ..., 5\}$. In world x, there are x recipients and x observers in the public sphere, and $5-x$ recipients and $5-x$ observers in the private sphere. Thus, the T0 world has no public sphere, the T5 world has no private sphere, while each of the other four worlds mixes public and private spheres in varying proportions. In order to be as close to the theory as possible, for each world, we let the dictator make binary decisions of contribution for its public and private spheres (if any) respectively. For all x, a prosocial action in the public sphere in world Tx costs the dictator $20x$ yuan, helps x recipients (each receiving 30 yuan), and is observed by all the five observers; a prosocial action in the private sphere in world Tx costs the dictator $20(5-x)$ yuan, helps $5-x$ recipients and is only observed by the $5-x$ observers in the private sphere. After the dictator makes the ten decisions for the six parallel worlds, one world is randomly selected (with equal probability) and the dictator's decisions for this world are sent to the five observers according to the above rule. The observers are then asked to simultaneously and independently evaluate the dictator's generosity in a scale between 0 and 5 based on their own observations. The sum of the five observers' evaluation is called the dictator's "generosity score" and will influence the dictator's income from the experiment.

This within-subject design has two advantages: First, it provides us with sufficient information about the dictator's strategy. Second, in this static experiment, the subjects have no opportunity to learn from repeated interactions, and thus may have no clue to decide if they are asked to make a one-shot decision in a single world (x). Our within-subject design, which presents six parallel worlds to the dictator and asks the dictator to make decisions for the six worlds simultaneously, helps the dictator understand the problem and think through it in its entirety. (See the translated experimental instructions

in the Online Appendix D.)

There are two between-subject treatments that vary in the stake of the "generosity score" for the dictator: "High Stake" treatment and "Low Stake" treatment. In the High Stake treatment, the dictator's income is determined by

$$100 \text{ yuan endowment} - 20 \times \#\text{recipients helped} + \text{generosity score} \times 5,$$

where "# recipients helped" means the number of recipients helped by the dictator in the realized world. In the Low Stake treatment, the dictator's income is determined by

$$100 \text{ yuan endowment} - 20 \times \#\text{recipients helped} + \text{generosity score} \times 3.$$

Put differently, given the scale of observers' evaluations in $[0, 5]$, the stake of the generosity score for the dictator's income is 125 yuan in the High Stake treatment, and 75 yuan in the Low Stake treatment, while the dictator's total endowment is 100 yuan. The between-subject design amounts to scaling up or down μ in the theoretical model.

Let $a^s_{Tx}, a^t_{Tx} \in \{0, 1\}$ denote the action in the private sphere and in the public sphere respectively in world Tx, for all $x \in \{0, 1, \ldots, 5\}$. Figure 1 and Proposition 2 predict that,

$$\underbrace{a^t_{T1} \geq a^t_{T2} \geq \cdots \geq a^t_{T5}}_{\text{cheap signaling}} = \underbrace{a^s_{T0} \geq a^s_{T1} \geq \cdots \geq a^s_{T4}}_{\text{moral licensing}}. \tag{11}$$

Given the binary nature of the actions, the theory predicts that a dictator will either donate all ($a^s_{Tx} = a^t_{Tx} = 1$ for all x), or never donate ($a^s_{Tx} = a^t_{Tx} = 0$ for all x), or choose to donate ($a = 1$) for the beginning decision(s) listed in Equation (11) and switch to $a = 0$ for the rest of the decisions. Thus, there is no much variation in a dictator's choices (with at most one switch point), if the dictator behaves according to the theory. In the experiment, if the subjects faced the same incentive scheme and their preference heterogeneity was small,[49] many subjects would behave similarly and the empirical results would hardly capture the potential diversity implied by Equation (11).[50]

We thus adopt the above-mentioned between-subject design by randomly assigning subjects to two treatments, with high or low stake. Given the random assignment of subjects into the two treatments, the High Stake treatment corresponds to a right-ward shift of the distribution of image concerns μ, relative to the Low Stake treatment. We expect this manipulation in the stake of generosity scores to generate more variations in the subjects' behavior.

[49]This is particularly the case when the subjects are from the same subject pool, e.g. students from the same university.

[50]This turns out to be true when we analyse the data from a single treatment, either the High Stake treatment or the Low Stake treatment. For example, in the Low Stake treatment, many subjects switch somewhere in the public sphere and so choose $a^s_{Tx} = 0$ for all x. Thus, it is hard to discern a decreasing trend in contribution in the private sphere when we look at this treatment separately. See Section 4.3 for details.

Equation (11) gives us the following hypotheses.

Hypothesis 1' (cheap signaling in the public sphere): For all $x \in \{1, 2, 3, 4\}$, $a^t_{Tx} \geq a^s_{T0} = a^t_{T5}$. That is, players are more likely to contribute in the public sphere of world T1, T2, T3 or T4 than in worlds T0 (the all-private baseline) and T5 (the all-public baseline).

Hypothesis 1' also implies that players are equally likely to contribute in worlds T0 and T5.

Hypothesis 2' (more expensive signaling in the public sphere): $a^t_{T1} \geq a^t_{T2} \geq a^t_{T3} \geq a^t_{T4}$. That is, players are less likely to contribute in the public sphere, as the public sphere expands.

Hypothesis 3' (moral licensing in the private sphere): For all $x \in \{1, 2, 3, 4\}$, $a^s_{Tx} \leq a^s_{T0} = a^t_{T5}$. That is, players are less likely to contribute in the private sphere of world T1, T2, T3 or T4 than in worlds T0 and T5.

Hypothesis 4' (augmented moral licensing in the private sphere): $a^s_{T1} \geq a^s_{T2} \geq a^s_{T3} \geq a^s_{T4}$. That is, players are less likely to contribute in the private sphere, as the public sphere expands.

Hypothesis 5' (misallocation in contributions): For all $x \in \{1, 2, 3, 4\}$, $a^t_{Tx} \geq a^s_{Tx}$. That is, in each world where the public sphere and private sphere coexist, players are more likely to contribute in the public sphere than in the private sphere.

The six worlds with varied t allow us to test the non-monotonic relationship between overall contributions and the size of the public sphere. Let overall contribution $\bar{a}$ be the sum of the numbers of the recipients helped in the public and private spheres, which ranges from 0 to 5.

Hypothesis 6 (non-monotonicity in total contributions): The sign of $d\bar{a}/dt$ is ambiguous. That is, the overall contribution changes non-monotonically as the public sphere expands.

Finally, we expect that the larger stake in the High Stake treatment will incentivize subjects to contribute more:

Hypothesis 7 (shift of image concerns): Subjects contribute more in the High Stake treatment than in the Low Stake treatment, other things equal.

4.2 Experimental implementation

We take advantage of a real-world charity fund in China in implementing our experiment. This charity fund is run by the Hebei Charitable Joint Foundation on the Alipay charity platform to help sanitation workers, and is officially certified. Low-income sanitation workers, especially those who do cleaning work in central business districts in big cities, usually cannot afford to buy proper lunches near their working area. The charity uses collected funds to finance nutritious lunches for groups of sanitation workers in China. In our experiment, we tell the subjects that we will donate the contributions collected from

the experiment to the charity in the name of Wuhan University, where the experiment is run. To help one recipient (sanitation worker), a subject donates 20 yuan (or equivalently USD2.7 at the time when the experiment was implemented), and the recipient receives 30 yuan (USD4.1), which is roughly the price of a nutritious lunch in big cities in China.

After the subjects make contribution decisions, we hire some student helpers, who do not participate in the experiment, as observers to evaluate the subjects' generosity according to the above-mentioned rule for a randomly selected world. It is common knowledge for the subjects that the observers receive a fixed payment, which is independent of the evaluations. The subjects' payments are determined after the evaluations are completed. They are paid in private via WeChat payment, which is a commonly used online payment method in China. We then make a lump-sum donation to the charity through the Alipay platform, according to the subjects' decisions in the randomly selected world. The subjects also receive their own generosity scores rated by the observers, as well as a soft copy of the receipt of the lump-sum donation (if they make any donation in the selected world). The dictators (subjects), recipients (sanitation workers) and observers are mutually anonymous to each other.

We ran the experiment at the economics laboratory of Wuhan University, China, in October 2022. In total, 179 subjects, with various majors, from the student subject pool of Wuhan University participated. The subjects had no previous experience in our experiment or similar experiments. 108 of them were assigned to the High Stake treatment and 71 in the Low Stake treatment. An experimental session lasted for about one hour. On average a subject received 100.5 yuan (or equivalently USD13.8) and the total amount of donation we made to the charity was 13,230 yuan.

4.3 Experimental results

The results from the static experiment support all the Hypotheses 1'-5', 6 and 7.

(a) Summary statistics

Behavior that is consistent with the theoretical predictions in Equation (11) can be divided into the following categories: (i) "Donate All": $a^t_{Tx} = a^s_{Tx} = 1$ for all $x \in \{0, 1, ..., 5\}$; (ii) "Keep All": $a^t_{Tx} = a^s_{Tx} = 0$ for all $x \in \{0, 1, ..., 5\}$; (iii) "Switch in Pub": There exists $1 \leq x^* \leq 4$ such that $a^t_{Tx} = 1$ for all $x \leq x^*$ and $a^t_{Tx} = 0$ otherwise, and $a^s_{Ty} = 0$ for all $y \in \{0, 1, 2, 3, 4\}$; (iv) "Switch in Pvt": There exists $0 \leq x^* \leq 3$ such that $a^s_{Tx} = 1$ for all $x \leq x^*$ and $a^s_{Tx} = 0$ otherwise, and $a^t_{Ty} = 1$ for all $y \in \{1, 2, 3, 4, 5\}$. Any behavior that does not belong to any of these categories is inconsistent with the theory and is called "Irrational behavior".

Table 3 shows the distribution of behavioral patterns for the whole sample and the individual treatments, respectively. Overall, the behavior of 134 out of the 179 subjects (74.8%) is consistent with the theory. The Fischer's exact test shows that there is no significant difference in the likelihood of consistent-to-the-theory behavior between the

High Stake treatment and the Low Stake treatment (p value =0.29). However, due to the larger stake from the "generosity score", the subjects in the High Stake treatment are more likely to choose $a^t_{Tx} = 1$ for all $x \in \{1, 2, 3, 4, 5\}$ (including "Donate All" and "Switch in Pvt") than in the Low Stake treatment (Fischer's exact test, p value < 0.001); meanwhile, the subjects in the Low Stake treatment are more likely to choose $a^s_{Tx} = 0$ for all $x \in \{0, 1, 2, 3, 4\}$ (including "Keep All" and "Switch in Pub") than in the High Stake treatment (Fischer's exact test, p value < 0.001).[51] These findings will bear some implications on the significance of results when we test some of the hypotheses with the individual treatments, as discussed below.

Table 3: Distribution of Behavioral Patterns

	Whole Sample		High Stake		Low Stake	
Behavioral Patterns	Num.of Subjects	Percent	Num.of Subjects	Percent	Num.of Subjects	Percent
Donate All	35	19.55	28	25.93	7	9.86
Switch in Pvt	45	25.14	39	36.11	6	8.45
Switch in Pub	38	21.23	12	11.11	26	36.62
Keep All	16	8.94	5	4.63	11	15.49
Irrational	45	25.14	24	22.22	21	29.58
Total	179	100	108	100	71	100

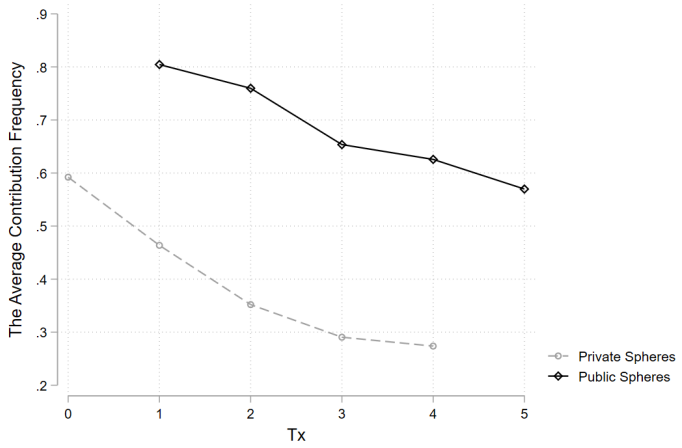
Figure 4 plots the average contribution frequency in public and private spheres in different worlds, with the whole sample presented at Figure 4(a), the High Stake treatment in Figure 4(b) and the Low Stake treatment in Figure 4(c). We have the following observations for the whole sample and for the individual (High Stake or Low Stake) treatments.

First, the average contribution frequency in world T0 is close to that in T5 ($\bar{a}^s_{T0} = 0.59 \approx \bar{a}^t_{T5} = 0.57$, and 171 out of the 179 subjects choose $a^s_{T0} = a^t_{T5}$).[52]

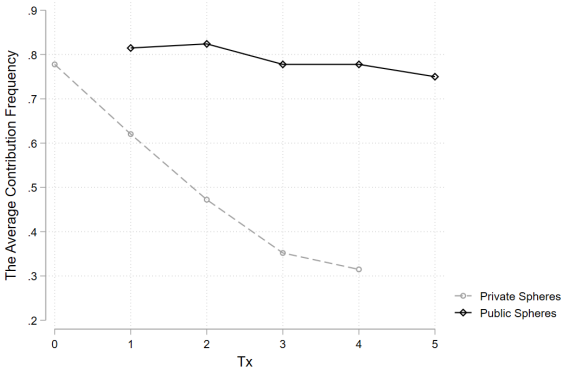
Second, in each of the worlds with the co-existence of the public and private spheres (i.e. T1, T2, T3 and T4), the average contribution frequency is higher in the public sphere than in the baseline case T5, and is lower in the private sphere than in the baseline case T0; that is, $\bar{a}^t_{Tx} > \bar{a}^t_{T5}$ and $\bar{a}^s_{Tx} < \bar{a}^s_{T0}$ for all $x \in \{1, 2, 3, 4\}$. If we look at the individual treatments separately, the pattern remains, and is stronger in the public sphere under the Low Stake treatment and in the private sphere under the High Stake treatment, but is weaker in the private sphere under the Low Stake treatment and in the public sphere

[51] In the High Stake treatment, 67 out of the 108 subjects choose $a^t_{Tx} = 1$ for all $x \in \{1, 2, 3, 4, 5\}$, while in the Low Stake treatment, only 13 out the 71 subjects do so. In the Low Stake treatment, 37 out of the 71 subjects choose $a^s_{Tx} = 0$ for all $x \in \{0, 1, 2, 3, 4\}$, while in the High Stake treatment, only 17 out of the 108 subjects do so.

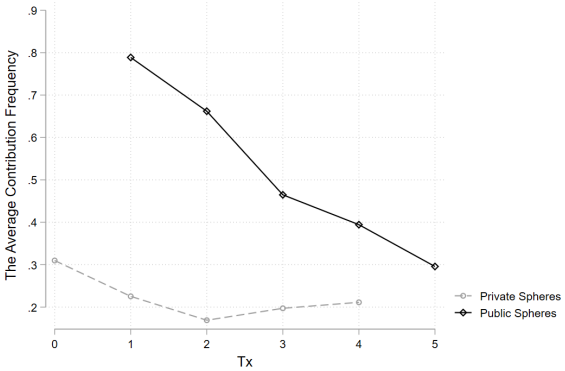
[52] Under the High Stake treatment, the average contribution frequency is 0.78 in world T0 and 0.75 in world T5. Under the Low Stake treatment, the average contribution frequency is 0.31 in world T0 and 0.30 in world T5. The McNemar's test shows that under each treatment (High Stake or Low Stake), there is no statistically significant difference in terms of contribution likelihood between world T0 and T5 (p values > 0.17). Table B3 in the Online Appendix reports the p values of all the McNemar's tests we conducted).



(a) Whole Sample



(b) High Stake



(c) Low Stake

Figure 4: The Average Contribution Frequency

under the High Stake treatment.[53] The McNemar's test shows that under the Low Stake treatment, contribution is significantly more likely for a^t_{Tx}, for all $x \in \{1, 2, 3, 4\}$, than for a^t_{T5}, and that under the High Stake treatment, contribution is significantly less likely for a^s_{Tx}, for all $x \in \{1, 2, 3, 4\}$, than for a^s_{T0} (p values < 0.08). Meanwhile, the test shows that under the Low Stake treatment, contribution is significantly less likely for a^s_{Tx}, for all $x \in \{1, ..., 4\}$, than for a^s_{T0} (p values < 0.1, except for a^s_{T4} where the p value $= 0.108$), and that under the High Stake treatment, contribution is significantly more likely for a^t_{T2} than for a^t_{T5} (p value < 0.1). Here, we compare a^s_{Tx} with a^s_{T0} and compare a^t_{Tx} with a^t_{T5} for all $x \in \{1, 2, 3, 4\}$. Since the average contribution frequency in world T0 is close to that in T5, the analysis comparing a^t_{Tx} with a^s_{T0} and comparing a^s_{Tx} with a^t_{T5} gives us qualitatively similar results. These findings, in general, lend support to Hypotheses 1'

[53]The weaker pattern there can be explained by the smaller behavioral variation in the public sphere of the Hight Stake treatment and in the private sphere of the Low Stake treatment. A majority (67 out of 108, or 62%) of the subjects under the High Stake treatment choose $a^t_{Tx} = 1$ for all $x \in \{1, 2, 3, 4, 5\}$, and a majority (37 out of 71, or 52.1%) of the subjects under the Low Stake treatment choose $a^s_{Tx} = 0$ for all $x \in \{0, 1, 2, 3, 4\}$.

and 3'.

Third, we observe that with the whole sample, the average contribution frequency decreases when the public sphere expands (moving from world T1 to T4), in both the public sphere and the private sphere. This is also the case for the private sphere under the High Stake treatment and the public sphere under the Low Stake treatment. These findings lend support to Hypotheses 2' and 4'. The regression analysis below will test the monotonicity results in Hypotheses 2'and 4' statistically.

Fourth, in each world where the public sphere and private sphere coexist, the average contribution frequency is higher in the public sphere than in the private sphere. That is, $\bar{a}^t_{Tx} > \bar{a}^s_{Tx}$ for all $x \in \{1, 2, 3, 4\}$. The McNemar's test shows that under each treatment (High Stake or Low Stake) and for each world $x \in \{1, 2, 3, 4\}$, players are significantly more likely to contribute in the public sphere than in the private sphere (p values < 0.01). These findings support Hypothesis 5'.

Figure 5 plots the overall contribution for the whole sample (solid line), the High Stake treatment (dashed line), and the Low Stake treatment (dotted line) respectively in different worlds. The relationship between the overall contribution and the magnitude of the public sphere ($x \in \{0, 1, 2, 3, 4, 5\}$) is indeed non-monotonic, which lends support to Hypothesis 6.[54]

Finally, the shift of image concerns leads to more contributions in the High Stake treatment than in the Low Stake treatment. This can be observed from Figure 4 and 5. The Mann-Whitney test conducted at the individual level shows that the overall contribution is significantly higher in the High Stake treatment than in the Low Stake treatment in each world (p values < 0.001), confirming Hypothesis 7.

(b) Regression results

Table 4 reports odds ratios from logit regressions of subjects' contribution decisions, clustering standard errors at the individual level. Column (1) controls for the size of the public sphere, $x \in \{0, 1, ..., 5\}$, which indicates the number of recipients in the public sphere, and a dummy variable *Public* indicating whether or not the decision is made in the public sphere. Column (2) further introduces the interaction term between x and *Public*. Column (1) and (2) control for individual fixed effects. Column (3) modifies the regression in Column (2), by replacing the individual fixed effects with a dummy variable *Low Stake* (indicating the Low Stake treatment) and deomographic variables (gender, age, major, year of study, and a dummy variable indicating previous experience of participation in behavioral experiments).

In Column (1), the odds ratio of *Public* is significantly greater than one. In Column (2)-(3), the odds ratios of $Public + Public \times x$ (by Wald test of linear restrictions) are also

[54]The theory predicts that $\bar{a}(t)$ is inverse U-shaped for uniform distributions. The relation indeed looks inverse U-shaped for the Low Stake treatment. The theory also predicts that $\bar{a}(t)$ is increasing when t is sufficiently low. However, the discrete nature of t in the experiment may not allow us to capture this property. We do not observe such a pattern from the High Stake treatment.

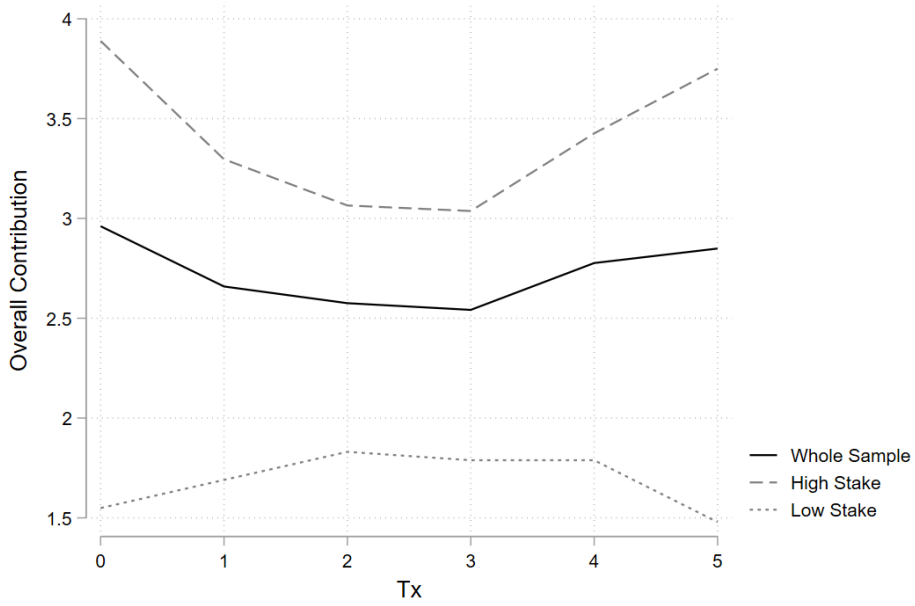


Figure 5: Overall Contribution

significantly greater than one. These findings support Hypothesis 5': Subjects are more likely to contribute in the public sphere than in the private sphere, other things equal. In both Column (2)-(3), the odds ratios of x and $x + Public \times x$ are all significantly smaller than one, suggesting that in both the private sphere and the public sphere, contribution is significantly decreasing in the magnitude of the public sphere $x \in \{0, 1, ..., 5\}$. There are two implications of this finding. First, when there is a co-existence of public and private spheres, contribution is significantly decreasing in the magnitude of the public sphere $x \in \{1, 2, 3, 4\}$, in both the private sphere and the public sphere, supporting Hypotheses 2' and 4'. Second, the finding also implies that contribution a^t_{Tx} for $x \in \{1, 2, 3, 4\}$ is higher than the baseline case a^t_{T5}, and contribution a^s_{Tx} for $x \in \{1, 2, 3, 4\}$ is lower than the baseline case a^s_{T0}, lending some support to Hypotheses 1' and 3'. Table 7 below, which uses the T0 world as the baseline, will more directly test Hypotheses 1' and 3'. Finally, in Column (3), the odds ratio of *Low Stake* is significantly smaller than one, which suggests lower contribution in the Low Stake treatment than in the High Stake treatment and confirms Hypothesis 7.

Column (1)-(3) of Table 4 use the whole sample in the regressions, treating the world T0 and T5 differently: $x = 0$ and $Public = 0$ for world T0 and $x = 5$ and $Public = 1$ for world T5. Column (3)-(5) repeat the regressions in Column (1)-(3) but exclude world T0 and T5. The results are qualitatively the same.

Table 5 repeats the regressions in Column (1)-(2) of Table 4 for the High Stake treatment (in its Column (1)-(2)) and the Low Stake treatment (in Column (3)-(4)) respectively. For both treatments, the results are qualitatively the same as those reported in

Table 4: Determinants of contribution behavior in the static experiment

	Whole Sample			T1–T4		
	(1)	(2)	(3)	(4)	(5)	(6)
x	0.443***	0.373***	0.671***	0.439***	0.424***	0.731***
	(0.043)	(0.062)	(0.038)	(0.063)	(0.089)	(0.050)
Public	53.124***	25.292***	5.006***	57.900***	49.011***	6.454***
	(22.589)	(15.590)	(1.379)	(26.603)	(42.085)	(2.143)
Public × x		1.371	1.076		1.070	0.960
		(0.299)	(0.100)		(0.322)	(0.114)
Low Stake			0.267***			0.328***
			(0.071)			(0.088)
Constant	32.206***	51.824***	0.008**	29.983***	33.318***	0.005**
	(8.320)	(24.206)	(0.018)	(12.389)	(20.669)	(0.012)
Wald Test of linear restrictions						
x + Public × x		0.512***	0.722***		0.453***	0.702***
Public + Public × x		34.665***	5.385***		52.450***	6.197***
Demographics	No	No	Yes	No	No	Yes
Individual Fixed Effects	Yes	Yes	No	Yes	Yes	No
Pseudo R-squared	0.617	0.620	0.183	0.651	0.651	0.187
N	1790	1790	1790	1432	1432	1432

Note: This table reports odds ratios from logit regressions, with standard errors clustered at the individual level. Variable $x \in \{0, 1, ..., 5\}$ indicates the size of public sphere. In Column (1)-(3), $x = 0$ and $Public = 0$ for world T0 and $x = 5$ and $Public = 1$ for world T5. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 4, except that in Column (2) the odd ratio of $x + Public \times x$ is smaller than one but statistically insignificant, while in Column (4) the odds ratio of x is smaller than one but statistically insignificant. These exceptions suggest that the decreasing contribution trend in x in the public sphere of the High Stake treatment and in the private sphere of the Low Stake treatment are not statistically significant. This can be explained by the lack of behavioral variation in these two scenarios (as mentioned, 62% of the subjects choose $a^t_{Tx} = 1$ for all $x \in \{1, 2, 3, 4, 5\}$ under the High Stake treatment and 52.1% choose $a^s_{Tx} = 0$ for all $x \in \{0, 1, 2, 3, 4\}$ under the Low Stake treatment). Each of the treatment captures some aspects of the behavioral diversity predicted by the theory.

Table 6 reports OLS regression results of subjects' overall contributions in a world, with robust standard errors clustered at the individual level. Column (1) controls for the size of the public sphere in the world, x. Column (2) further introduces the quadratic term of x. Column (1) and (2) control for individual fixed effects. Column (3)-(4) modify the regressions in Column (1)-(2), by replacing the individual fixed effects with variable *Low*

Table 5: Determinants of contribution behavior in the static experiment (by treatment)

	High Stake		Low Stake	
	(1)	(2)	(3)	(4)
x	0.445***	0.223***	0.441***	0.735
	(0.060)	(0.065)	(0.060)	(0.148)
Public	53.312***	2.884	52.874***	523.870***
	(31.923)	(2.217)	(30.850)	(593.013)
Public × x		3.538***		0.387***
		(1.160)		(0.137)
Constant	31.811***	246.119***	0.107***	0.036***
	(11.384)	(224.502)	(0.027)	(0.018)
Wald Test of linear restrictions				
x + Public × x		0.790		0.285***
Public + Public × x		10.204***		202.915***
Individual Fixed Effects	Yes	Yes	Yes	Yes
Pseudo R-squared	0.604	0.640	0.583	0.606
N	1080	1080	710	710

Note: This table reports odds ratios from logit regressions, with standard errors clustered at the individual level. Variable $x \in \{0, 1, ..., 5\}$ indicates the size of public sphere. $x = 0$ and $Public = 0$ for world T0 and $x = 5$ and $Public = 1$ for world T5. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Stake and deomographic variables. Column (5)-(8) repeat the regressions in Column (1)-(2) for the High Stake treatment and for the Low Stake treatment respectively. We find that, when only the linear term of x is added (Column (1), (3), (5) and (7)), the coefficients of x are statistically insigificant and the magnitudes are small. When the quadratic term of x is added, then both the linear term and the quadratic term become statistically significant for the whole sample (Column (2) and (4)) and the High Stake treatment (Column (6)), but are still insignificant for the Low Stake treatment (Column (8)). These results confirm the non-monotonic relationship between the overall contribution and the size of the public sphere (Hypothesis 6). Meanwhile, the coefficients of *Low Stake* are negative and statistically significant, confirming Hypothesis 7.

The above regressions look at the effect of public sphere expansion by controlling for the variable x which indicates the size of the public sphere. Table 7, which also reports odds ratios from logit regressions of subjects' contribution decisions, takes a different approach. In the experiment, a subject makes binary decisions in 10 scenarios in total, which differ in whether the decision is made in the public or private sphere and in the size of the public sphere. Table 7 uses the world T0 as the baseline and introduces dichotomous variables that indicate the other nine scenarios $T1Pub$,..., $T4Pub$, $T1Pvt$,...,

Table 6: Determinants of overall contributions in the static experiment

	Whole Sample				High Stake		Low Stake	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
x	-0.007	-0.288**	-0.007	-0.288**	-0.010	-0.640***	-0.003	0.249
	(0.017)	(0.133)	(0.016)	(0.122)	(0.022)	(0.154)	(0.028)	(0.224)
x^2		0.056**		0.056**		0.126***		-0.050
		(0.027)		(0.024)		(0.031)		(0.045)
Low Stake			-1.662***	-1.662***				
			(0.274)	(0.274)				
Constant	4.850***	5.038***	-1.925	-1.738	4.857***	5.278***	0.507***	0.339**
	(0.043)	(0.096)	(2.297)	(2.297)	(0.054)	(0.113)	(0.071)	(0.161)
Demographics	No	No	Yes	Yes	No	No	No	No
Individual Fixed Effects	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Adjusted R-squared	0.685	0.689	0.181	0.184	0.644	0.673	0.607	0.611
N	1074	1074	1074	1074	648	648	426	426

Note: This table reports OLS regression results of subjects' overall contributions (varying between 0 and 5), with robust standard errors clustered at individual level. Variable x indicates the size of public sphere. *** p <0.01, ** p <0.05, * p <0.1.

$T4Pvt$, and $T5$. Such an approach facilitates direct pairwise comparison between any two of the scenarios and allows us to test Hypotheses 1' and 3' more directly. Column (1) controls for individual fixed effects, column (2) replaces the individual fixed effects by variable *Low Stake* and demographic variables, while column (3) and (4) repeat the regression in column (1) with the subsample of the High Stake treatment and Low Stake treatment respectively. In all the columns, standard errors are clustered at the individual level.

In Column (1)-(2) of Table 7, the odds ratios of $T5$ are statistically insignificant, confirming that there is no significant difference in terms of contribution frequency between world T0 and T5. The odds ratios of $T1Pub$, $T2Pub$, and $T3Pub$ are statistically significant and greater than one (the odds ratios of $T4Pub$ are greater than one but statistically insignificant). These results lend support to Hypothesis 1'. The odds ratios of $T1Pvt$, $T2Pvt$, $T3Pvt$ and $T4Pvt$ are all statistically significant and smaller than one, which supports Hypothesis 3'. By Wald test of linear restrictions, the odds ratios of $T2Pub-T1Pub$, $T3Pub - T2Pub$, $T4Pub - T3Pub$, $T5 - T4Pub$ are smaller than one and in particular, those of $T3Pub - T2Pub$ and $T5 - T4Pub$ are statistically significant; the odds ratios of $T2Pvt - T1Pvt$, $T3Pvt - T2Pvt$, $T4Pvt - T3Pvt$ are statistically significant and smaller than one (except for $T4Pvt - T3Pvt$ whose odds ratios are smaller than one but statistically insignificant). These results lend support to Hypotheses 2' and 4'. Finally, the odds ratios of $T1Pub - T1Pvt$, $T2Pub - T2Pvt$, $T3Pub - T3Pvt$, and $T4Pub - T4Pvt$ are all statistically significant and greater than one, supporting Hypothesis 5'. In Column (3)-(4) with only one of the treatments, the results are qualitatively similar but some

become less statistically significant, due to the fact that a majority of the subjects under the High Stake treatment choose $a^t_{Tx} = 1$ for all $x \in \{1, 2, 3, 4, 5\}$, and a majority of the subjects under the Low Stake treatment choose $a^s_{Tx} = 0$ for all $x \in \{0, 1, 2, 3, 4\}$.

5 Conclusion

With tech giants' and governments' rapid deployment of data technologies, individual behavior becomes more transparent in many aspects of life. Optimists argue that transparency will promote socially valued behavior. Our paper challenges this "conventional wisdom" both theoretically and empirically. Agents' social interactions involve multitasking between their public and private spheres. Because behavior in the public sphere is more widely observable, the agents behave more prosocially in the public sphere than in the private sphere. However an increase in transparency generates crowding out in both, public and private, spheres: When the public sphere expands, public sphere signaling is no longer cheap, reducing prosociality in the public sphere. This reduced prosociality in turn augments moral licensing in the private sphere, which crowds out prosociality in that sphere; so agents behave less prosocially in both spheres. Overall, since the substitution effect and the level effect work in opposite directions, the aggregate effect of public sphere expansion on prosociality is ambiguous.

We design two experiments to test the theory. We find strong evidence that, consistent with the premises of the theory, reputational concerns drive behavior. In the dynamic experiment, we identify distinct channels for why this is so: direct reciprocity ("I am nice to you as you have been nice to me"), indirect reciprocity ("I am nice to you as you have been nice to others") and social norm obedience ("I am nice to you as people are nice overall"). In the static experiment, indirect reciprocity is the channel of reputational concerns. More importantly, the findings from both experiments support the key theoretical predictions of our model. (1) The cheap-signaling effect generates a higher prosociality in the public sphere than in the all-private treatment. (2) Prosociality in the public sphere decreases with the size of the public sphere. (3) The moral-licensing effect generates a lower prosociality in the private sphere than in the all-private treatment. (4) Prosociality in the private sphere decreases with the size of the public sphere. (5) The subjects misallocate efforts by behaving more prosocially in the public sphere than in the private sphere. (6) Overall prosociality may not increase monotonically with an expansion of the public sphere.

Table 7: Determinants of contribution behavior in the static experiment (Dichotomous scenario variables)

	Whole Sample		High Stake	Low Stake
	(1)	(2)	(3)	(4)
T1Pub	11.557***	3.276***	1.668	188.546***
	(6.496)	(0.830)	(1.189)	(225.026)
T2Pub	6.330***	2.439***	1.911	29.499***
	(2.847)	(0.519)	(1.150)	(23.460)
T3Pub	1.908*	1.353*	1.000	4.457***
	(0.678)	(0.226)	(0.522)	(2.551)
T4Pub	1.420	1.177	1.000	2.312
	(0.420)	(0.163)	(0.426)	(1.193)
T5	0.794	0.899	0.700	0.861
	(0.128)	(0.067)	(0.181)	(0.226)
T1Pvt	0.269***	0.549***	0.163***	0.374*
	(0.085)	(0.078)	(0.076)	(0.207)
T2Pvt	0.079***	0.323***	0.035***	0.154**
	(0.031)	(0.055)	(0.020)	(0.114)
T3Pvt	0.034***	0.235***	0.007***	0.250*
	(0.018)	(0.048)	(0.006)	(0.187)
T4Pvt	0.025***	0.214***	0.003***	0.308
	(0.016)	(0.047)	(0.004)	(0.228)
Low Stake		0.265***		
		(0.071)		
Constant	61.589***	0.009**	269.920***	0.055***
	(26.475)	(0.020)	(226.179)	(0.028)
Wald Test of linear restrictions				
T1Pub - T1Pvt	42.982***	5.963***	10.224***	503.788***
T2Pub - T2Pvt	79.652***	7.550***	54.148***	191.665***
T3Pub - T3Pvt	56.227***	5.755***	139.063***	17.816***
T4Pub - T4Pvt	55.897***	5.488***	295.093***	7.498***
T2Pub - T1Pub	0.548	0.744	1.146	0.156**
T3Pub - T2Pub	0.301***	0.555***	0.523	0.151***
T4Pub - T3Pub	0.744	0.870	1.000	0.519*
T5 - T4Pub	0.559**	0.764**	0.700	0.372*
T2Pvt - T1Pvt	0.296***	0.588***	0.216***	0.411
T3Pvt - T2Pvt	0.427**	0.728**	0.204***	1.625
T4Pvt - T3Pvt	0.748	0.912	0.471	1.232
Demographics	No	Yes	No	No
Individual FE	Yes	No	Yes	Yes
Pseudo R-squared	0.623	0.186	0.642	0.617
N	1790	1790	1080	710

Note: This table reports odds ratios from logit regressions, with standard errors clustered at the individual level. *** p <0.01, ** p <0.05, * p <0.1.

References

Acquisti, A., Taylor, C., and L. Wagman (2016), "The Economics of Privacy," *Journal of Economic Literature*, 54: 442–492.

Algan, Y., Benkler, Y., Fuster-Morell, M. and J. Hergueux (2016), "Cooperation in a Peer Production Economy: Experimental Evidence from Wikipedia," Sciences Po Working Paper, November.

Ali, S.N., and R. Bénabou (2020), "Image Versus Information: Changing Societal Norms and Optimal Privacy," *American Economic Journal: Microeconomics*, 12(3): 116-164.

Ariely, D., Bracha, A. and S. Meier (2009), "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially," *American Economic Review*, 99(1): 544–555.

Ashraf, N. and O. Bandiera (2018), "Social Incentives in Organizations?" *Annual Review of Economics*, 10: 439–463.

Ashraf, N., Bandiera, O. and J. Kelsey (2014), "No Margin, No Mission? A Field Experiment on Incentives for Public Services Delivery," *Journal of Public Economics* 120: 1–17.

Austen-Smith, D., and R. Fryer (2005), "An Economic Analysis of 'Acting White'," *Quarterly Journal of Economics*, 120(2): 551–583.

Bénabou, R., and J. Tirole (2006), "Incentives and Prosocial Behavior," *American Economic Review*, 96(5): 1652–1678.

Bénabou, R., and J. Tirole (2011a), "Identity, Morals and Taboos: Beliefs as Assets," *Quarterly Journal of Economics*, 126(2): 805–855.

Bénabou, R., and J. Tirole (2011b), "Laws and Norms," Technical report, National Bureau of Economic Research.

Bernheim, D., and A. Bodoh-Creed (2019), "Pervasive Signaling," mimeo.

Besley, T., Jensen, A. and T. Persson (2015), "Norms, Enforcement, and Tax Evasion," CEPR Discussion Paper No DP10372.

Bolton, G. E., Katok, E., and A. Ockenfels (2005), "Cooperation among Strangers with Limited Information about Reputation," *Journal of Public Economics*, 89(8): 1457–1468.

Bursztyn, L., and R. Jensen (2017), "Social Image and Economic Behavior in the Field: Identifying, Understanding and Shaping Social Pressure," *Annual Review of Economics*, 9: 131–153.

Bursztyn, L., Egorov. G. and R. Jensen (2019) "Cool to Be Smart or Smart to Be Cool? Understanding Peer Pressure in Education," *Review of Economic Studies*, 86(4): 1487-1526 (2019).

Bursztyn, L., González, A. L., and Yanagizawa-Drott, D. (2020a). "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia," *American Economic Review*, 110(10), 2997-3029.

Bursztyn, L., Egorov, G., and S. Fiorin (2020b), "From Extreme to Mainstream: The Erosion of Social Norms," *American Economic Review*, 110(11): 3522–48.

Chen, D. (2016), "The Deterrent Effect of the Death Penalty? Evidence from British Commutations During World War I," TSE Working Paper no. 16-706, R&R *American Economic Review*.

Daughety, A., and J. Reinganum (2010), "Public Goods, Social Pressure, and the Choice between Privacy and Publicity," *American Economic Journal: Microeconomics*, 2(2): 191–221.

DellaVigna, S., List, J., Malmendier, U. and G. Rao (2017), "Voting to Tell Others," *Review of Economic Studies*, 84: 143–181.

Dewatripont, M., Jewitt, I. and J. Tirole (1999), "The Economics of Career Concerns, I: Comparing Information Structure," *Review of Economic Studies*, 66(1): 183–198.

Effron, D., Cameron, J., and B. Monin (2009) "Endorsing Obama Licenses Favoring Whites," *Journal of Experimental Social Psychology*, 45: 590-593.

Effron, D., Miller, D., and B. Monin (2012) "Inventing Racist Roads Not Taken: The Licensing Effect of Immoral Counterfactual Behaviors," *Journal of Personality and Social Psychology*, 103(6): 916-932.

Engelmann, D., and U. Fischbacher (2009), "Indirect Reciprocity and Strategic Reputation Building in an Experimental Helping Game," *Games and Economic Behavior*, 67(2): 399–407.

Fischbacher, U. (2007), "z-Tree: Zurich Toolbox for Ready-made Economic Experiments," *Experimental Economics*, 10(2): 171–178.

Freeman, R. (1997), "Working for Nothing: The Supply of Volunteer Labor," *Journal of Labor Economics*, 15(1): S140–66.

Funk, P. (2010), "Social Incentives and Voter Turnout: Evidence from the Swiss Mail Ballot System," *Journal of the European Economic Association*, 8(5): 1077–1103.

Galbiati, R., Schlag K. and J. van der Weele (2013), "Sanctions that Signal: An Experiment," *Journal of Economic Behavior and Organization*, 94: 34–51.

Gerber A., Green, D. and C. Larimer (2008), "Social Pressure and Voter Turnout: Evidence from a Large- Scale Field Experiment," *American Political Science Review*, 102(1): 33–48.

Holmström, B., and P. Milgrom (1991), "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization*, 7: 24–52.

Hirshleifer, J. (1971), "The Private and Social Value of Information and the Reward to Inventive Activity," *American Economic Review*, 61: 561–574.

Jewitt, I. (2004), "Notes on the Shape of Distributions," unpublished.

Jia, R., and T. Persson (2017), "Individual vs. Social Motives in Identity Choice: Theory and Evidence from China," mimeo.

Karing, A. (2019), "Social Signaling and Childhood Immunization: A Field Experiment in Sierra Leone," mimeo, UC Berkeley.

Lacetera, N., Macis, M. and R. Slonim (2012), "Will There Be Blood? Incentives and Displacement Effects in Pro-social Behavior," *American Economic Journal: Economic Policy*, 4(1): 186–223.

Lanier, J. (2010), "Missing Persons," In *You Are Not a Gadget,* London: Penguin.

Merritt, A., Effron, D., Fein, S., Savitsky, K., Tuller, D., and B. Monin, B. (2012) "The Strategic Pursuit of Moral Credentials," *Journal of Experimental Social Psychology*, 48: 774-777.

Milinski, M., Semmann, D., Bakker, T. C., and H.-J. Krambeck (2001), "Cooperation through Indirect Reciprocity: Image Scoring or Standing Strategy?," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1484): 2495–2501.

Monin, B. and D. Miller (2001) "Moral Credentials and the Expression of Prejudice," *Journal of Personality and Social Psychology,* 81(1): 33-43.

Nowak, M. A. and K. Sigmund (1998), "The Dynamics of Indirect Reciprocity," *Journal of Theoretical Biology*, 194(4): 561–574.

Nowak, M. A. and K. Sigmund (2005), "Evolution of Indirect Reciprocity," *Nature*, 437(7063): 1291–1298.

Perez-Truglia, R. and G. Cruces (2017), "Partisan Interactions: Evidence from a Field Experiment in the United States," *Journal of Political Economy*, 125(4): 1208–1243.

Russell, S. (2019), *Human Compatible: Artificial Intelligence and the Problem of Control*, Viking, Penguin Random House.

Seinen, I. and A. Schram (2006), "Social Status and Group Norms: Indirect Reciprocity in a Repeated Helping Experiment," *European economic review*, 50(3): 581–602.

Smith, A. (1759), *The Theory of Moral Sentiments*, Strand & Edinburgh: A. Millar.

Tirole, J. (2021), "Digital Dystopia," *American Economic Review*, 111(6): 2007–2048.

Tirole, J. (2022), "Safe Spaces: Shelters or Tribes?," mimeo.

Watanabe, T., Takezawa, M., Y., N., A., K., H., Y., Nakamura, M., Miyashita, Y., and A. Masuda (2014), "Two Distinct Neural Mechanisms Underlying Indirect Reciprocity," PNAS, 111(11): 3990–95.

Wedekind, C. and M. Milinski (2000), "Cooperation through Image Scoring in Humans," *Science*, 288(5467): 850–852.

Williams, R. (2012). "Using the Margins Command to Estimate and Interpret Adjusted Predictions and Marginal Effects," *The Stata Journal*, 12(2), 308–331.

Appendix: Theory

(a) *Off-the-equilibrium-path beliefs.* Let $\bar{a}^t \in [0,1]$ denote the average contribution in the public sphere. Under a deterministic and symmetric behavior, $\bar{a}^t \in \{0,1\}$. Suppose for conciseness that $v^s < \sup v$, so both $a_{ij} = 0$ and $a_{ij} = 1$ are on the equilibrium path for (ij) in the private sphere. If $v^t = 0$, specify that $\hat{v}_{ji} = 0$ if $\bar{a}^t < 1$ where $\hat{v}_{ji}$ is j's posterior estimate of v_i. If $v^t > 0$, then both $\bar{a}^t = 0$ and $\bar{a}^t = 1$ are on-path behaviors. For $\bar{a}^t < 1$, set $\hat{v}_{ji} = M^-(v^t)$ (which, incidentally, covers the case $v^t = 0$) no matter whether the (ij) relationship is in the public or private sphere. These beliefs sustain the deterministic and symmetric behavior described by (5) and (6) as an equilibrium.

(b) *Uniqueness of the deterministic symmetric equilibrium.*

Lemma 1 *A sufficient condition for there to always be more contributions in the public sphere ($v^s \geq v^t$) in any equilibrium is that the density of the type distribution be non-increasing (e.g. uniform).*

Proof of Lemma 1. Suppose to the contrary that $v^s < v^t$ and let $M(v_0, v_1)$ denote the mean of v over the interval $[v_0, v_1)$.

Behavior in the private sphere, being unobservable except to the counterparty, does not impact the reputation in the public sphere. So, for any $v_i \in [v^s, v^t)$,

$$s\left[(v_i - c) + \mu[M(v^s, v^t) - M^-(v^s)]\right] \geq 0.$$

Similarly the fact that in this interval, agents do not want to contribute publicly implies that:

$$t(v_i - c) + \mu\left[s[M^+(v^t) - M(v^s, v^t)] + t[M^+(v^t) - M^-(v^t)]\right] \leq 0.$$

These two inequalities are inconsistent if

$$M(v^s, v^t) - M^-(v^s) < [M^+(v^t) - M^-(v^t)] + \frac{s}{t}[M^+(v^t) - M(v^s, v^t)].$$

The latter condition is satisfied in particular (for $s > 0$) if for $v^s < v^t$

$$M^+(v^t) - M^-(v^t) \geq M(v^s, v^t) - M^-(v^s). \tag{12}$$

Inequality (12) is satisfied at $v^s = v^t$ (since $M^+(v^t) \geq v^t$). Furthermore, applying Jewitt (2004)'s lemma on $[0, v^t]$, $M(v^s, v^t) - M^-(v^s)$ is non-decreasing if the density f is non-increasing. ■

Finally, one cannot guarantee that (6) has a unique solution, unless $f' \leq 0$ and so the additional term is non-increasing with prosocial behavior in the public sphere. Nonetheless, it can be shown that the prosocial behavior in the public and private spheres

is decreasing in t at stable equilibria.

(c) *Uniform distribution.* Normalize $e = 1$ and suppose that $v \sim U[0,1]$. Then $\bar{a}(0) = \bar{a}(1) = 1 - (c - \frac{\mu}{2})$. To match the assumptions in the text, posit that

$$c > 1 \text{ (image concerns are needed for the provision of the public good)}$$

and

$$0 < v^* = c - \frac{\mu}{2} < 1 \text{ (interior solution when behavior is publicly observed).}$$

Let

$$0 < t_0 \equiv \frac{1}{1 + \frac{2}{\mu}} < t_1 \equiv \min\left\{\frac{\mu^2}{4(c-1)(\frac{\mu}{2}+1)}, 1\right\}.$$

Then

$$\bar{a}(t) = \begin{cases} 1 - (1-t)(c - \frac{\mu}{2}) & \text{for} \quad t \leq t_0 \\ 1 + \frac{1 - \frac{2c}{\mu}}{\frac{2}{\mu} - \frac{\mu}{2} + \frac{\mu}{2t}} & \text{for} \quad t_0 \leq t \leq t_1 \\ \frac{\mu}{2} - t(c-1) & \text{for} \quad t \geq t_1 \text{ (if } t_1 < 1). \end{cases}$$

So $\bar{a}(t)$ increases linearly with t in the first region, decreases in the second region (is convex if $\mu < 2$, concave if $\theta > 2$), and decreases linearly in the third region (if it exists).

Appendix: Discussion of Table 1

(a) *Addressing missing values in Table 1.* In Table 1, column (1) includes only $T40Pub$, $T40Pvt$, $T80Pub$, $T80Pvt$ as the right-hand side variables. Starting from column (2), $\bar{a}_{i\tau}$ and $(\bar{a}_{ji\tau} - \bar{a}_{i\tau})$ are included. Since variable $\bar{a}_{i\tau}$ (contribution frequency of previous partners to subject i before round τ) is not well defined for the first-round interaction, the data of the first round interaction is dropped for columns (2)-(5). Variable $\bar{a}_{ji\tau}$ (contribution frequency of subject j to subject i before round τ) is not well defined if the two players in the current round have never met before. Column (2) dropped the observations with missing values, while starting from column (3), we use $\bar{a}_{i\tau}$ to replace the missing values of $\bar{a}_{ji\tau}$, and so $(\bar{a}_{ji\tau} - \bar{a}_{i\tau})$ takes the value of 0. Here, we assume that when two players interact for the first time, subject i assumes that subject j is no different from the previous partners with whom she has interacted before. Column (4) and (5) further include social scores $\bar{s}_{i\tau}$ and $(s_{j\tau} - \bar{s}_{i\tau})$. We address the missing social scores in treatment T0 in the following way. Note that, in T40 and T80, $\bar{s}_{i\tau}$ is used to update belief on the average "goodness of society". Column (4) uses treatment T0's cumulative contribution frequency (from round 1 to round $\tau - 1$) to replace the missing value of $\bar{s}_{i\tau}$, while column (5) uses treatment T0's overall contribution frequency to replace the missing value of $\bar{s}_{i\tau}$. The former is time-varying while the latter is time-invariant. Variable $(s_{j\tau} - \bar{s}_{i\tau})$ takes the value of 0 for treatment T0.

(b) *Adding $\bar{a}_{i\tau}$ and $(\bar{a}_{ji\tau} - \bar{a}_{i\tau})$ overturns results in Table 1.* In Table 1, the odds ratios of $T40Pub - T80Pub$ and $T40Pvt - T80Pvt$ are overturned from being smaller than one to being greater one when we add $\bar{a}_{i\tau}$ and $(\bar{a}_{ji\tau} - \bar{a}_{i\tau})$ to the regression (column (2)-(3)). The odds ratios of $\bar{a}_{i\tau}$ are much larger than one and statistically significant, implying that subjects were more likely to contribute if they were well treated by previous partners; meanwhile, we find that $\bar{a}_{i\tau}$ is significantly and positively associated with $T80Pub$ and $T80Pvt$ and significantly and negatively associated with $T40Pub$ and $T40Pvt$ (two-sample t test; p-values<0.001).[55] Therefore, omitting $\bar{a}_{i\tau}$ in the regression would overestimate the effects of $T80Pub$ and $T80Pvt$, and underestimate the effects of $T40Pub$ and $T40Pvt$, and so underestimate the odds ratios of $T40Pub - T80Pub$ and $T40Pvt - T80Pvt$, as in column (1).

The fact that $\bar{a}_{i\tau}$ is positively associated with $T80Pub$ and $T80Pvt$ indicates that $\bar{a}_{i\tau}$ tends to be higher in treatment T80. Apparently, $\bar{a}_{i\tau}$ is affected by the distribution of public and private spheres in the first $\tau - 1$ rounds. (Note that "Pub"/"Pvt" in "T80Pub" or "T80Pvt" indicates the regime in round τ and has nothing to do with the regime distribution in the first $\tau - 1$ rounds.) In our experiment, it turns out that all the first 10 rounds in treatment T80 were in the public sphere. Therefore, subjects in treatment T80 tend to start with a higher $\bar{a}_{i\tau}$, given that subjects contribute more in public than in private (Hypothesis 5). The higher $\bar{a}_{i\tau}$ in earlier rounds then induced more contributions in subsequent rounds (strategic complementarity, as shown by the odds ratio of $\bar{a}_{i\tau}$). This constitutes positive feedback and explains the higher $\bar{a}_{i\tau}$ under treatment T80.

[55] $(\bar{a}_{ji\tau} - \bar{a}_{i\tau})$ is not significantly associated with any of these treatment-regime variables though.

Online Appendix A: Figures

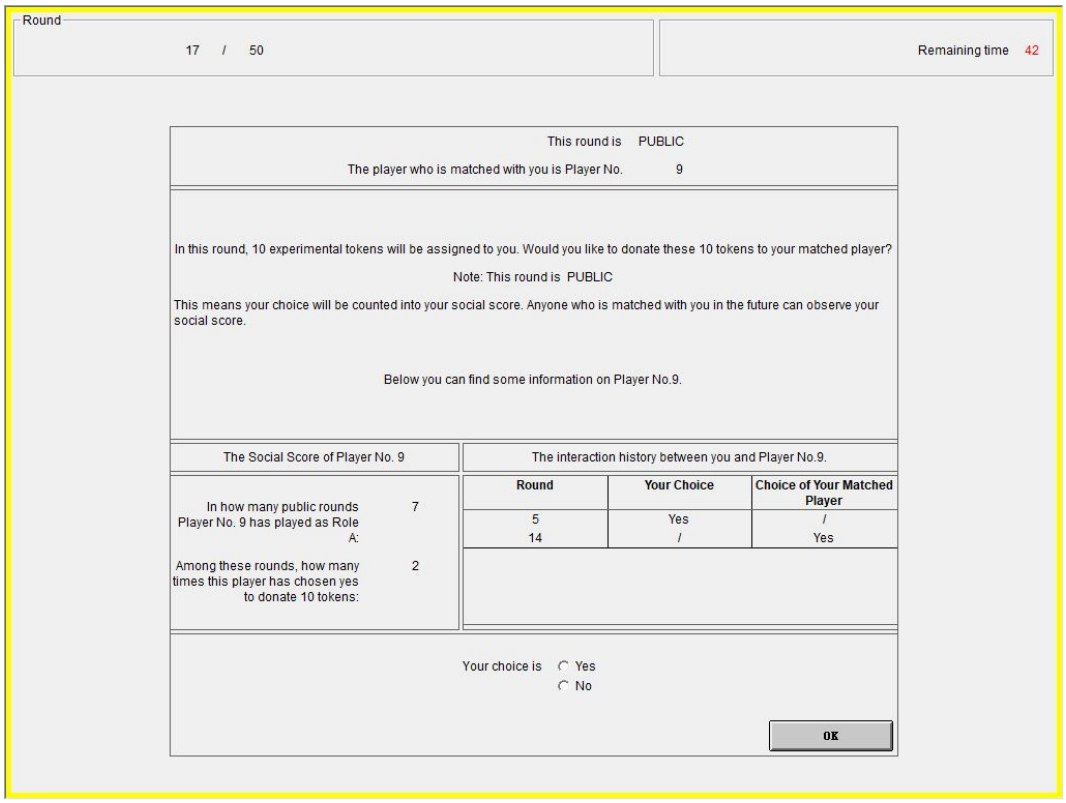


Figure A1: Screenshot of a player A's decision interface in treatment T40/T80 (translated in English)

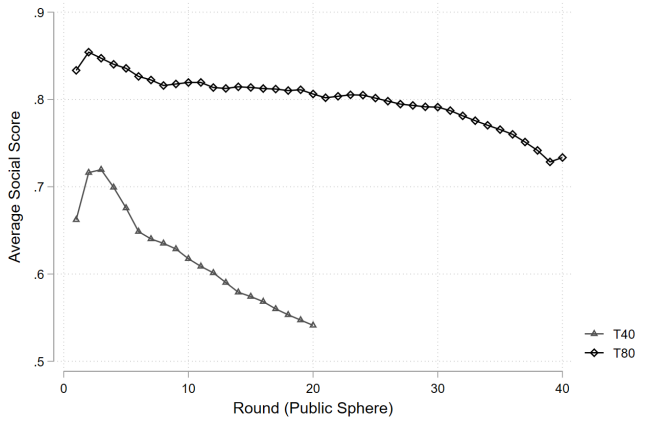


Figure A2: Time trend of social scores

Notes: This figure shows the time trend of average social scores in T40 and T80. The social score is defined as a subject's cumulative contribution frequency in the public sphere. The horizontal line represents the sequence of rounds in the public sphere; e.g. round 10 means the tenth round in the public sphere.

Online Appendix B: Tables

Table B1: Determinants of contribution behavior in the dynamic experiment (All treatment groups): SE clustered at individual level

	(1)	(2)	(3)	(4)	(5)
T40*Pub*	1.975***	3.653***	2.889***	2.430***	2.221***
	(0.266)	(0.665)	(0.417)	(0.401)	(0.392)
T40*Pvt*	0.476***	0.736	0.534***	0.373***	0.343***
	(0.064)	(0.145)	(0.079)	(0.070)	(0.069)
T80*Pub*	5.120***	1.605**	2.486***	1.595**	1.525**
	(0.692)	(0.324)	(0.363)	(0.301)	(0.314)
T80*Pvt*	1.251	0.321***	0.521***	0.297***	0.284***
	(0.198)	(0.077)	(0.096)	(0.066)	(0.067)
$\bar{a}_{i\tau}$		204.431***	20.816***	29.832***	26.892***
		(69.629)	(3.968)	(6.249)	(5.493)
$\bar{s}_{i\tau}$				2.791***	2.832***
				(0.931)	(0.930)
$(\bar{a}_{ji\tau} - \bar{a}_{i\tau})$		22.598***	19.228***	15.531***	15.439***
		(2.844)	(2.282)	(1.867)	(1.852)
$(s_{j\tau} - \bar{s}_{i\tau})$				9.884***	9.785***
				(1.794)	(1.767)
Wald Test of linear restrictions					
T40*Pub*-T40*Pvt*	4.153***	4.966***	5.410***	6.521***	6.478***
T80*Pub*-T80*Pvt*	4.091***	4.995***	4.773***	5.369***	5.379***
T40*Pub*-T80*Pub*	0.386***	2.276***	1.162	1.524**	1.456**
T40*Pvt*-T80*Pvt*	0.380***	2.289***	1.025	1.255	1.209
Demographics	Yes	Yes	Yes	Yes	Yes
Round Fixed Effects	Yes	Yes	Yes	Yes	Yes
Pseudo R-squared	0.165	0.400	0.297	0.326	0.323
N	10917	5640	10698	10477	10550
Replacement of Missing Values	No	No	Yes	Yes	Yes

Note: This table reports odds ratios from logit regressions of contribution behavior. Demographic characteristics of the decision maker and round fixed effects are controlled for. Standard errors clustered at the individual level are reported in parentheses. *, **, and *** represent significance at 10, 5, and 1% levels, respectively. In column (2), observations with missing values of $\bar{a}_{i\tau}$ or $\bar{a}_{ji\tau}$ are dropped. In columns (3)-(5), we use the values of $\bar{a}_{i\tau}$ to replace the missing values of $\bar{a}_{ji\tau}$. In column (4), we fill in the missing values of $\bar{s}_{i\tau}$ in treatment T0 with the treatment mean of cumulative contribution frequency before the current round; in column (5), the missing values of $\bar{s}_{i\tau}$ in treatment T0 are replaced by the treatment mean of contribution frequency including all rounds. The variable $(s_{j\tau} - \bar{s}_{i\tau})$ takes value 0 in columns (4)-(5) for treatment T0.

Table B2: Determinants of contribution behavior in the dynamic experiment (Treatment T40 and T80): SE clustered at individual level

	(1)	(2)	(3)
Pub	5.562***	6.574***	5.955***
	(1.213)	(1.038)	(0.922)
T80	0.411***	0.746	0.130***
	(0.108)	(0.164)	(0.076)
Pub × T80	1.113	0.919	1.083
	(0.316)	(0.201)	(0.239)
$\bar{a}_{i\tau}$	276.195***	30.782***	31.307***
	(121.457)	(8.258)	(8.418)
$\bar{s}_{i\tau}$	0.585	3.225***	1.406
	(0.341)	(1.114)	(0.559)
$(\bar{a}_{ji\tau} - \bar{a}_{i\tau})$	15.413***	12.853***	12.435***
	(2.564)	(2.030)	(1.962)
$(s_{j\tau} - \bar{s}_{i\tau})$	5.733***	10.508***	5.604***
	(1.404)	(1.906)	(1.010)
$\bar{s}_{i\tau}$× T80			8.814***
			(6.265)
$(s_{j\tau} - \bar{s}_{i\tau})$× T80			6.544***
			(2.553)
Wald Test of linear restrictions			
Pub + Pub × T80	6.189***	6.044***	6.446***
T80 + Pub × T80	0.457***	0.686**	0.141***
Demographics	Yes	Yes	Yes
Round Fixed Effects	Yes	Yes	Yes
Pseudo R-squared	0.423	0.363	0.369
N	3689	6973	6973
Replacement of Missing Values	No	Yes	Yes

Note: This table reports odds ratios from logit regressions of contribution behavior including data from treatment T40 and T80. Demographic characteristics of the decision maker and round fixed effects are controlled for. Standard errors clustered at the individual level are reported in parentheses. *, **, and *** represent significance at 10, 5, and 1% levels, respectively. In column (1), we omit the observations if there are missing values. In columns (2) and (3), we use the values of variable $\bar{a}_{i\tau}$ to replace the missing values of variable $\bar{a}_{ji\tau}$.

Table B3: p values of McNemar's tests on contribution in the static experiment (by treatment)

	High Stake	Low Stake
T1Pub vs T1Pvt	0.0006	0.0000
T2Pub vs T2Pvt	0.0000	0.0000
T3Pub vs T3Pvt	0.0000	0.0001
T4Pub vs T4Pvt	0.0000	0.0067
T0 vs T5	0.1797	0.5637
T1Pub vs T5	0.1779	0.0000
T2Pub vs T5	0.0881	0.0000
T3Pub vs T5	0.4669	0.0073
T4Pub vs T5	0.3657	0.0707
T1Pvt vs T0	0.0002	0.0833
T2Pvt vs T0	0.0000	0.0124
T3Pvt vs T0	0.0000	0.0593
T4Pvt vs T0	0.0000	0.1083
T1Pub vs T0	0.4652	0.0000
T2Pub vs T0	0.2752	0.0000
T3Pub vs T0	1.0000	0.0116
T4Pub vs T0	1.0000	0.1088
T1Pvt vs T5	0.0010	0.1655
T2Pvt vs T5	0.0000	0.0389
T3Pvt vs T5	0.0000	0.1083
T4Pvt vs T5	0.0000	0.1573

Online Appendix C: English Translation of Experimental Instructions for Dynamic Experiment (Treatment T80)

INSTRUCTIONS

Welcome to the experiment. This experiment studies decision-making among individuals. The experiment is expected to last no longer than one hour and 20 minutes. Please read the following instructions carefully; the payment you will obtain at the end of the experiment depends both on your decisions and the decisions made by others. At the end of today's session, you will be paid privately.

During the experiment, your identity will not be disclosed, and your decisions will not be associated with your identity. In order to ensure smooth implementation of the experiment, please do not leave the laboratory until the end of the experiment. During the experiment, please turn off your electronic devices; communication with any other participant is not allowed.

Your task

In each round of the experiment, you will be randomly matched with another player in this room. You will take turns playing the following two roles:

Role A (Decision Maker): At the beginning of the round, you will get 10 units of experimental tokens, and the computer will ask you if you are willing to donate these 10 tokens. If your choice is "yes", the player who is matched with you (Role B) will receive 15 tokens. If your choice is "No", you will keep the 10 tokens as your earning in this round.

Role B (Recipient): You will observe the choice of your matched player (who is playing Role A in this round). You do not take any action in this round.

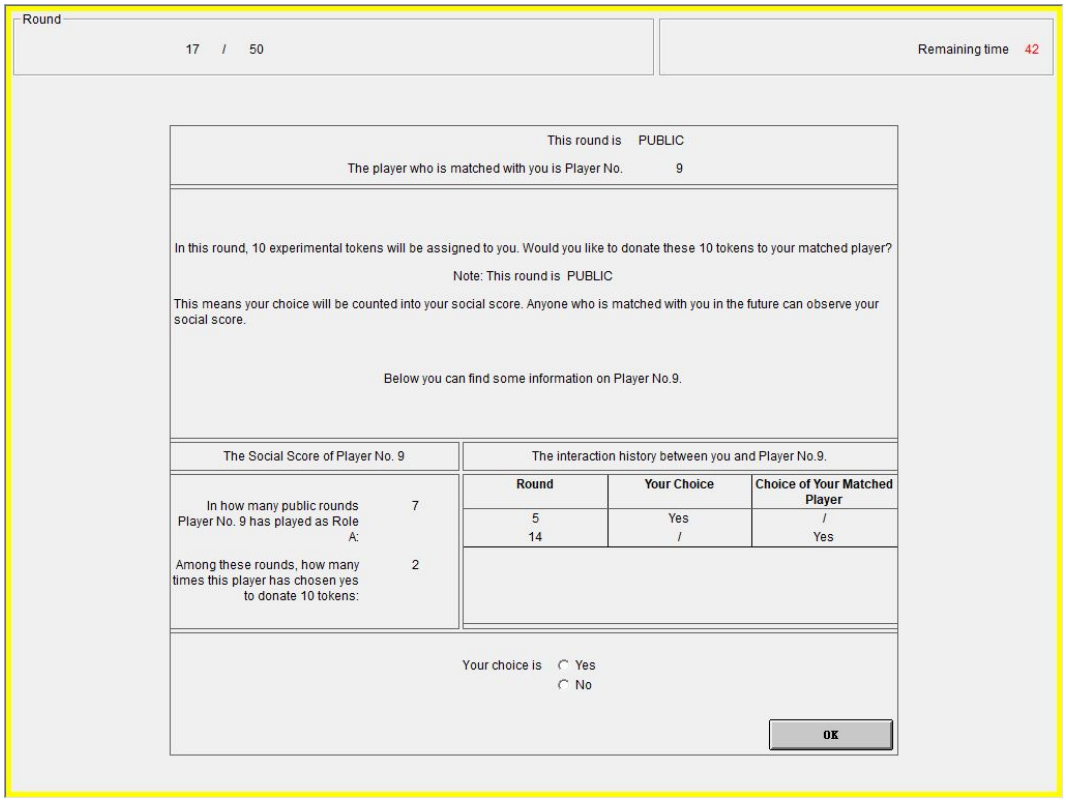
There will be 50 rounds in this experiment. You will play the two roles in turns. That is, you will play Role A in 25 (either odd or even) rounds and play Role B in the other 25 rounds.

Each round will be randomly assigned to one of two possible interactive environments, namely **PUBLIC** and **PRIVATE**: In the **PUBLIC** environment, if you play Role A, your choice will be recorded publicly. To be specific, anyone who is matched with you in the future can observe "in how many public rounds you have played Role A", and among these rounds, "how many times you have chosen yes to donate 10 tokens". This constitutes your "social score". In the **PRIVATE** environment, if you play Role A, your choice will only be observed by the player who is matched with you in the current round, and will not be recorded in your "social score", and so players matched with you in the future will not observe your choice in this round.

In this experiment, a total of **80%** of the rounds are public.

Given the above settings, when you play Role A and are asked to make a decision, you can see the following information about the player who is matched with you (as shown in

the picture below):



- **The history of interactions between you two.** Although we randomly rematch players in every round, you may have repeated interactions with one particular other player. The history of the previous interactions between you two will be shown, including the roles played and the choices of donation made.
- **Role B's "social score"**. That is, in how many previous rounds your matched player has played Role A in the public environment, and in how many of these rounds, your matched player has chosen "Yes".

Your earning

In each round your earning will be determined in the following ways:

Role A:

- If you choose to donate the 10 experimental tokens, your earning of this round is 0;
- If you choose not to donate the 10 experimental tokens, your earning of this round is 10;

Role B:

- If your matched player (Role A) chooses "yes" to donate, your earning of this round is 15;
- If your matched player chooses not to donate, your earning for this round is 0.

After completing all the rounds, the computer will show you your earning of each round in a table. For example:

rounds	earning
1	10
2	15
3	0
4	0
5	0
6	15
7	10
8	0
.....	

The computer will then randomly select 4 out of the 50 rounds for your payment. Your final payment will be equal to the sum of your earnings in the 4 selected rounds plus a show up fee of 30 yuan. The exchange rate is set to be one experimental token being equivalent to one yuan. For example, if you earn 25 experimental tokens in the four selected rounds, your final payment is $30 + 25 \times 1 = 55$ yuan.

The Rundown of the Experiment

1. At the beginning of each round, the computer will randomly pair you with another participant.
2. The computer will tell you your role in this round and whether it is PUBLIC or PRIVATE in this round.
3. If your role is A, the computer will ask: "Are you willing to donate 10 experimental tokens to the player who is matched with you?" Please answer this question by clicking "Yes" or "No". Before you make the choice, some information about your matched player (the interaction history between you two and the matched player's social score) will be displayed on the screen.
4. If your role is B, after your matched player makes the decision you will see his/her choice and some information of your matched player on the screen (the interaction

history between you two and the matched player's social score). You then need to click "OK" to enter the next round.

5. Steps 1 to 4 will be repeated in the next round until all the 50 rounds are completed.

6. After all the rounds are completed, the computer will display your earning per round on the screen.

7. Please complete a questionnaire on the computer. The content of the questionnaire will not be associated to your identity.

Administration

Please use the mouse in front of you to enter your decision. Your decisions and your monetary earning will be kept confidential. After completing the experiment, you will get your payment. We will ask you to sign your earning to acknowledge your receipt of the payment. You are then free to leave. Now you can start. Good luck!

Online Appendix D: English Translation of Experimental Instructions for Static Experiment (Low Stake treatment)

INSTRUCTIONS

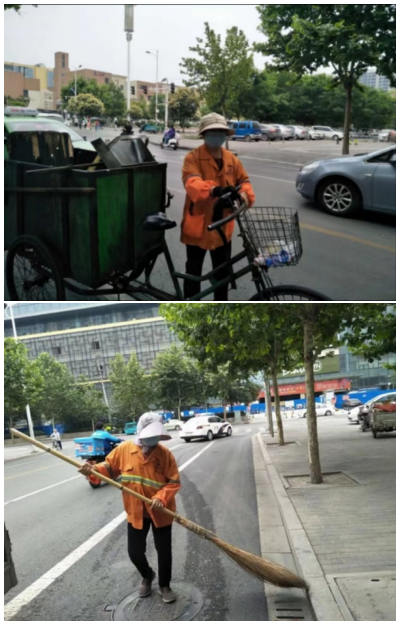
Welcome to the experiment. Please read the instructions carefully; the payment you will obtain after the experiment depends both on your decisions and the decisions made by others. The experiment contains two stages. Stage I will be conducted today. It will take no longer than one hour. After Stage II is finished, we will pay you privately through WeChat transfer.

During the experiment, your identity will not be disclosed, and your decisions will not be associated with your identity. In order to ensure smooth implementation of the experiment, please do not leave the laboratory until the end of Stage I. During the experiment, please turn off your electronic devices.

Stage I

A Charity Fund

The following information is important for the decisions you will make. Our experiment is related to a charity fund on the Alipay charity platform that aims to help sanitation workers.



Description of Charity Fund

The city's "beauticians", early morning runners. A group of people are wearing orange cleaning uniforms, holding a broom in their hands, or riding a cleaning tricycle. Regardless of season and weather, they appear in the streets and alleys every morning on time, silently cleaning the environment for us.

Auntie Wang is 65 years old. Her son touched a high-voltage line at a very young age, resulting in incapacitation. The 40 years old son has been taken care of by Auntie Wang since that. She said: "People in our village know that we are poor, and introduced this job to me. I have been doing this job for over 20 years. Every morning I start to go to the working area around three o'clock. Although it is very hard, I treasure this job since my family relies on my salary. So I save as much as possible, and my lunch is also brought from home." The volunteers saw that Auntie Wang's lunchbox was filled with a cold steamed bun and leftovers from the last night.

There are many sanitation workers like Auntie Wang, who often eat cold food brought from home because they can't afford to spend money on lunch. But in the summer, the food may spoil as the weather gets hot. Thus, having lunch becomes a big problem for them.

"Going out with the sun and dew, wearing the stars and moon to return." With hardworking hands, the sanitation workers do their best to clean the environment and beautify the city. To thank them for their efforts, the "One Lunch Warms a City" charity fund is launched to provide sanitation workers with nutritious and delicious lunches.

Your task

You now have RMB100 as your endowment.

You will decide, under different circumstances, whether to help some sanitation workers by providing a nutritious lunch for them. If you choose to help, your cost of helping each worker is RMB20, and the worker will receive RMB30. You can also choose not to help.

There are 5 potential **"recipients"** (sanitation workers). Their role is passive. We will transfer the donations generated from the experiment in the name of Wuhan University after the experiment. Meanwhile, there are 5 **"observers"** who will observe some of your choices and will make decisions that influence your final income based on their observations in Stage II.

The 5 "observers" are anonymous students at Wuhan University.

The recipients are in either an "X" or a "Y" scenario: If the recipient is in the "X" scenario, your choice towards him/her will be observed by all observers. If the recipient is in the "Y" scenario, your choice towards him/her will only be observed by some observers (explained below).

Meanwhile, there are 6 possible parallel worlds: T0: All 5 recipients are in the Y-scenario.

T1: There is 1 recipient in the X-scenario and 4 recipients in the Y-scenario.

T2: There are 2 recipients in the X-scenario and 3 recipients in the Y-scenario.

T3: There are 3 recipients in the X-scenario and 2 recipients in the Y-scenario.

T4: There are 4 recipients in the X-scenario and 1 recipient in the Y-scenario.

T5: All 5 recipients are in the X-scenario.

You don't need to specify which recipient you are going to help; you only need to tell us about your decisions under different circumstances, i.e., in a parallel world T0, T1, T2, T3, T4, or T5, 1) in the X-scenario, whether you choose to help or not; 2) in the Y-scenario, whether you choose to help or not.

In the parallel world T0: all 5 recipients are in the Y-scenario.
If you choose to help, you will pay 100 yuan (to help 5 recipients), and your choice of helping will be observed by the 5 observers; if you choose not to help, you do not need to pay the 100 yuan, and your choice of not helping will also be observed by the 5 observers.

In the parallel world T1: There is 1 recipient in the X-scenario and 4 recipients in the Y-scenario.
T1 (X-scenario): If you choose to help, you will pay 20 yuan (to help 1 recipient), and your choice of helping will be observed by the 5 observers; if you choose not to help, you do not need to pay the 20 yuan, and your choice of not helping will also be observed by the 5 observers.

T1 (Y-scenario): If you choose to help, you will pay 80 yuan (to help 4 recipients), and your choice of helping will be observed by 4 observers; if you choose not to help, you do not need to pay the 80 yuan, and your choice of not helping will also be observed by these 4 observers.

In the parallel world T2: There are 2 recipients in the X-scenario and 3 recipients in the Y-scenario.
T2 (X-scenario): If you choose to help, you will pay 40 yuan (to help 2 recipients), and your choice of helping will be observed by the 5 observers; if you choose not to help, you do not need to pay the 40 yuan, and your choice of not helping will also be observed by the 5 observers.

T2 (Y-scenario): If you choose to help, you will pay 60 yuan (to help 3 recipients), and your choice of helping will be observed by 3 observers; if you choose not to help, you do not need to pay the 60 yuan, and your choice of not helping will also be observed by these 3 observers.

In the parallel world T3: There are 3 recipients in the X-scenario and 2 recipients in the Y-scenario.
T3 (X-scenario): If you choose to help, you will pay 60 yuan (to help 3 recipients), and your choice of helping will be observed by the 5 observers; if you choose not to help, you do not need to pay the 60 yuan, and your choice of not helping will also be observed by

the 5 observers.

T3 (Y-scenario): If you choose to help, you will pay 40 yuan (to help 2 recipients), and your choice of helping will be observed by 2 observers; if you choose not to help, you do not need to pay the 40 yuan, and your choice of not helping will also be observed by these 2 observers.

In the parallel world T4: There are 4 recipients in the X-scenario and 1 recipient in the Y-scenario.
T4 (X-scenario): If you choose to help, you will pay 80 yuan (to help 4 recipients), and your choice of helping will be observed by the 5 observers; if you choose not to help, you do not need to pay the 80 yuan, and your choice of not helping will also be observed by the 5 observers.

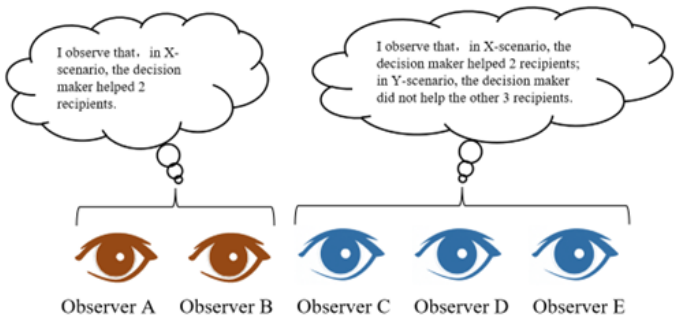
T4 (Y-scenario): If you choose to help, you will pay 20 yuan (to help 1 recipient), and your choice of helping will be observed by 1 observer; if you choose not to help, you do not need to pay the 20 yuan, and your choice of not helping will also be observed by the 1 observer.

In the parallel world T5: all 5 recipients are in the X-scenario.
If you choose to help, you will pay 100 yuan (to help 5 recipients), and your choice of helping will be observed by the 5 observers; if you choose not to help, you do not need to pay the 100 yuan, and your choice of not helping will also be observed by the 5 observers.

Phase II

We will randomly select a parallel world with equal probability from T0/T1/T2/T3/T4 /T5 and send your choices in this world to the 5 observers according to the rule described above. The decisions you make in the other parallel worlds will not be known to the observers.

Taking the T2 world as an example, if you choose to help in the X-scenario and not to help in the Y-scenario, the figure below shows the information observed by the 5 observers: 2 observers (brown eyes) only know that you help the two recipients in the X-scenario;



meanwhile, the other 3 observers (blue eyes) not only know that you help two recipients in X-scenario, but they also know that you do not help 3 other recipients in Y-scenario.

After the observers observe (some of) your choices, each observer will rate your generosity in a scale of 0 to 5. Note: The observers' ratings on you do not affect their own income. Their income is fixed. Each of the 5 observers makes their decisions independently. There will be no communication between them.

Your Earning

Your payment = 100 yuan (endowment) - 20 yuan × number of recipients you help + (sum of the ratings from 5 observers) × 3

Stage II will be completed within 10 days. After that, we will make the donation generated to the charity program. We will also pay you for the experiment via WeChat transfer and inform you of the ratings from the observers within 10 days.

The observers will not see your name or any other identification information. Your decisions, income and ratings from the observers will not be disclosed to anyone else.

An Illustration

An illustration is provided below.

Suppose your decisions are as follows:

	T0	T1	T2	T3	T4	T5
X-scenario		☒	☒	☒	☒	☒
Y-scenario	☒	☒	☐	☐	☐	

☒ denotes "to help", ☐ denotes "not to help"

Suppose the parallel world T3 is drawn, and your decisions in this world are to help in the X-scenario, but not in the Y-scenario.

Among the 5 observers, 3 of them only observe that you choose to help 3 recipients in the X-scenario. The remaining 2 observers not only observe that you choose to help in the X-scenario, but they can also see that you choose not to help two other recipients in the Y-scenario.

Suppose that the 3 observers who can only see your choice in the X-scenario give you ratings of U, V, and W respectively, and the 2 observers who can see your choices in both X-scenario and Y-scenario give you ratings M and N, respectively. U, V, W, M and N all lie between 0 and 5. Your total payment will be: 100 – 60 + (U+V+W+M+N) × 3 yuan.

A recap

You will have a donation decision to make vis-a-vis each of 5 potential recipients. In parallel world Tz, where $z \in \{0, 1, 2, 3, 4, 5\}$, your choice (the same) for z recipients in the X-scenario will be observed by all 5 observers; your choice (the same) for the $5 - z$ other recipients in the Y-scenario will be observed by only $5 - z$ observers.

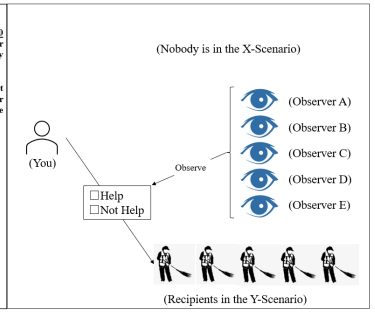
Please tick your choice in each graph:

The Parallel World: T0

T0

If you choose to help, you will pay 100 yuan (to help 5 recipients), and your choice of helping will be observed by the 5 observers.

If you choose not to help, you do not need to pay the 100 yuan, and your choice of not helping will also be observed by the 5 observers.



The Parallel World: T1

T1 (X)

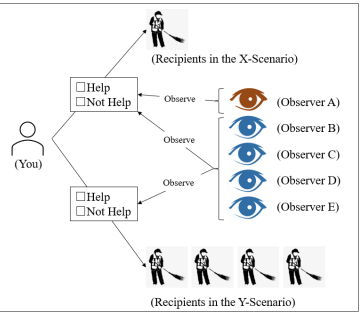
If you choose to help, you will pay 20 yuan (to help 1 recipient), and your choice of helping will be observed by the 5 observers.

If you choose not to help, you do not need to pay the 20 yuan, and your choice of not helping will also be observed by the 5 observers.

T1 (Y)

If you choose to help, you will pay 80 yuan (to help 4 recipients), and your choice of helping will be observed by 4 observers.

If you choose not to help, you do not need to pay the 80 yuan, and your choice of not helping will also be observed by these 4 observers.



The Parallel World: T2

T2 (X)

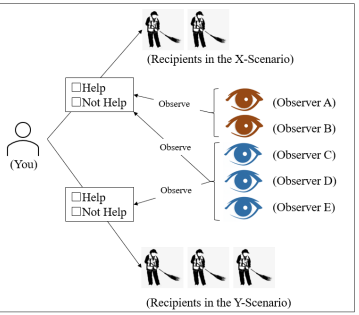
If you choose to help, you will pay 40 yuan (to help 2 recipients), and your choice of helping will be observed by the 5 observers.

If you choose not to help, you do not need to pay the 40 yuan, and your choice of not helping will also be observed by the 5 observers.

T2 (Y)

If you choose to help, you will pay 60 yuan (to help 3 recipients), and your choice of helping will be observed by 3 observers.

If you choose not to help, you do not need to pay the 60 yuan, and your choice of not helping will also be observed by these 3 observers.



The Parallel World: T3

T3 (X)

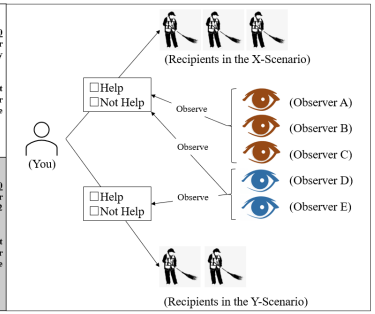
If you choose to help, you will pay 60 yuan (to help 3 recipients), and your choice of helping will be observed by the 5 observers.

If you choose not to help, you do not need to pay the 60 yuan, and your choice of not helping will also be observed by the 5 observers.

T3 (Y)

If you choose to help, you will pay 40 yuan (to help 2 recipients), and your choice of helping will be observed by 2 observers.

If you choose not to help, you do not need to pay the 40 yuan, and your choice of not helping will also be observed by these 2 observers.



The Parallel World: T4

T4 (X)

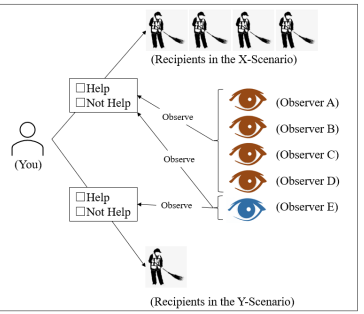
If you choose to help, you will pay 80 yuan (to help 4 recipients), and your choice of helping will be observed by the 5 observers.

If you choose not to help, you do not need to pay the 80 yuan, and your choice of not helping will also be observed by the 5 observers.

T4 (Y)

If you choose to help, you will pay 20 yuan (to help 1 recipient), and your choice of helping will be observed by 1 observer.

If you choose not to help, you do not need to pay the 20 yuan, and your choice of not helping will also be observed by the 1 observer.

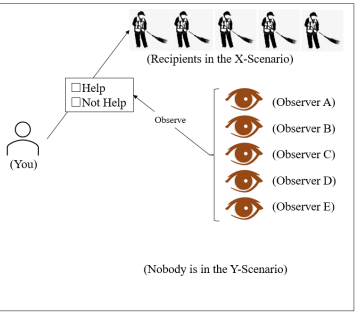


The Parallel World: T5

T5

If you choose to help, you will pay 100 yuan (to help 5 recipients), and your choice of helping will be observed by the 5 observers.

If you choose not to help, you do not need to pay the 100 yuan, and your choice of not helping will also be observed by the 5 observers.



We will randomly select one of the above six parallel worlds with equal probability and reveal your choices in this parallel world to the observers according to the rule described above; the observers will then decide on ratings, and your payment will be finalized. You will receive your payment via WeChat transfer together with the ratings from the anonymous observers and a copy of the electronic receipt on the donations to the charity (if any).