Crowding-out induced by signalling: Incentives, reputation and gender

Juan B. González

Supervised by Nicolas Jacquemet

June 1, 2022

M1 Economics and Psychology

Contents

Abstract

Why do monetary incentives crowd prosocial behaviour out and often backfire? This study experimentally shows that: (1) Crowding-out of prosociality is induced by reputation signalling. When decisions are public and reputation is at stake, incentives crowd out intrinsic motivation. Thus, monetary incentives are more likely to be counterproductive for public prosocial actions than for private ones. (2) There are gender differences in crowding-out, as incentives crowd prosociality out more for women than for men. (3) Second-order beliefs explain most of these gender differences.

JEL Classification: *D64, J16, L31, Z13.*

Acknowledgements

I want to thank Nicolas Jacquemet for his guidance through all the research process, and Pablo Brañas for his comments and support. Huge thanks also to my friends for their advice and care, especially to Arvid and Fanni for improving this research and Marta for always being there.

1 Introduction

Picture the following situation. A daycare is determined to reduce the number of parents that, violating the social norm, arrive late to pick up their children. They decide to charge late-coming parents a small fine, to give them an incentive to behave "well". Would the policy work? No need to imagine, Gneezy and Rustichini (2000) conducted an experiment to test it. To the surprise of many economists, the fine not only didn't reduce the number of late pick-ups, it actually resulted in more parents arriving late, even after the daycare withdrew the fine (Gneezy and Rustichini, 2000). Possibly without reading that result, Norway introduced a fine to penalize longer hospital stays. Again, the incentive had the opposite effect, increasing the length of hospital stays (Holmås et al., 2010). Research finds similar counterintuitive effects of incentives in other domains, such as charitable giving (Ariely et al., 2009; DellaVigna et al., 2013; Exley, 2016; Lazear et al., 2012), blood donations (Lacetera and Macis, 2010a; Goette and Stutzer, 2020; Mellström and Johannesson, 2008) and other prosocial actions (Fehr and Rockenbach, 2003; Henrich et al., 2010; Fryer, 2011).

The opposite than expected effect of incentives on the target behaviour is known as *crowding-out*, and has been extensively researched due to its policy implications for incentive design [1]. Despite all this research, findings are still inconclusive and sometimes contradictory. Besides, certain questions remained unanswered: although gender differences in crowding-out have been identified, the mechanisms behind them remain unknown. To address these issues, this study aims to answer three main questions: (1) Is crowding-out induced by signalling? (2) Are there gender differences in crowding-out? (3) In affirmative case, what are the

[1] See Fehr and Falk (2002); Gneezy et al. (2011); Bowles and Polania-Reyes (2012) and Kamenica (2012) for reviews on the topic.

mechanisms behind these gender differences?

To answer these questions, I replicate and extend the study conducted by Ariely et al. (2009). Using a 2x2 mixed experimental design (N = 178), this study shows that: (1) crowding-out is induced by reputation signalling, as incentives are less effective when offered in public than in private; (2) women crowd out significantly more of prosocial behaviour than men; (3) the difference across gender is mainly due to differences in second-order prosociality beliefs. Overall, this study offers supporting evidence for crowding-out induced by signalling, as well as novel findings on whether and why women crowd out more than men. The paper proceeds as follows. Section 2 presents the theoretical background and the hypotheses that arise from it, Section 3 the experimental design and data collection and Section 4 the results. Section 5 discusses the implications of these findings and concludes.

2 Why do people crowd out?

How can incentives hinder the target behaviour? Let's go back to the example of the daycare: why did parents arrive later when they had to pay a fine? One explanation is that when the daycare set the fine, parents interpreted it as a message or *signal*: the fine was the price of the misdemeanor. This signal changed the parents' mentality from social or moral (doing the right thing) to monetary (willingness to pay). A complementary explanation is that, before the fine, parents that arrived on time were trying to build their reputation by sending a *signal out* to others: they were good and prosocial people, doing the right thing. However, when the fine was introduced, arriving on time didn't send a clear prosocial signal

anymore. Others could interpret their action just as cheapness: those in time may only want to avoid paying the fine. Henceforth, we will refer to these two explanations as *signalling-in* and *signalling-out*, respectively.

2.1 A model of crowding-out

2.1.1 Signalling-in

Incentives provide an extrinsic motivation that is directly internalized by the decision maker. But incentives are also signals that affect decision-making indirectly by changing the situation and information ecology. Incentives can be interpreted as cues about the situation and the right response to it, as well as about the motives and expectations of the person administering the rewards. They thus provide crucial situational cues that frame the decision and affect the outcomes (Tversky and Kahneman, 1981; Ross and Nisbett, 2011).

The signal incentives provide can cause a complete shift in the perceived situation. In his pioneering book *The Gift Relationship*, Titmuss (1970) considered that incentives for prosocial actions might induce a "market mentality" instead of a social one [2]. This mentality shift, also called "moral disengagement" (Bandura, 1991), has been experimentally proven in diverse settings (Henrich et al., 2010; Heyman and Ariely, 2004; Fehr and Rockenbach, 2003). Experimental evidence is supported by neural imagery. The presence of incentives in a social decision (the Trust Game [3]) reduces activity in areas associated with social valuation

[2]When Titmuss (1970) introduced the idea that offering incentives for prosocial activities could backfire, it was received with harsh criticism from prominent economists (Arrow, 1972; McLean and Poulton, 1986). In contrast, the idea was welcomed in the field of psychology, where researchers found many instances of incentives crowding out intrinsic motivation (Deci et al., 1975, 1999; Lepper et al., 1982).

[3]In the Trust Game, Player 1 decides how much money to send to Player 2, who receives a multiple of the amount. Then Player 2 decides how much of the multiplied amount to send back to Player 1. While Player

(ventromedial prefrontal cortex and orbitofrontal cortex) while it increases activity in the parietal cortex, usually implicated in profit maximization (Li et al., 2009; Fehr and Krajbich, 2013). The shift in neural activity from social valuation to profit maximization has behavioral consequences: it reduces prosocial decisions, inducing crowding-out.

Therefore, incentives act as signals that change the framing of the decision. In other words, while prosocial behaviour is often motivated by social preferences, incentives can *crowd* the social motivation *out* and induce a monetary mindset. The model developed by Bowles and Polania-Reyes (2012) captures how incentives signal to the agent and reduce the intrinsic value of the prosocial action:

$$v_a = \lambda_0(1 + \phi(y)\lambda_c + y\lambda_m) \tag{1}$$

In their model, the intrinsic value of the prosocial action v_a is explained by the individual's social preferences λ_0. However, the presence $\phi(y)$ and magnitude y of incentives can reduce the intrinsic value of the action if $\lambda_c < 0$ and $\lambda_m < 0$.

2.1.2 Signalling-out

Incentives don't only offer a signal to the agent, they also change the signal the agent sends to others when behaving prosocially. Prosocial behaviour can be caused not only by the warm-glow of intrinsic motivation (Andreoni, 1990) or social preferences (Fehr and Schmidt, 1999), also by reputational concerns. Being perceived by others as prosocial has a beneficial impact on the agent through direct or indirect reciprocity and the avoidance of

1's decision is a risky investment choice, Player 2's is solely a social dilemma.

social punishment [4] (Fehr and Gächter, 2000; Kurzban et al., 2015; Roberts et al., 2021).

Therefore, doing a good deed is a way of signalling a prosocial reputational and individuals derive utility from being considered as prosocial (Andreoni and Bernheim, 2009). In this setting, incentives act as a confounder, increasing the noise of the signal. As such, they make less likely the attribution of a prosocial activity to intrinsic motivations (which build reputation) and more likely the attribution to extrinsic motivations (which could harm reputation). The presence of incentives can tell the agent that others won't consider prosocial behaviour as purely intrinsically motivated. Thus, public incentives can reduce the reputational signalling-out or even inverse its sign (Bénabou and Tirole, 2006). Indeed, experimental evidence proves that crowding-out is induced by reputation signalling: public incentives crowd-out intrinsic motivations more than private ones (Ariely et al., 2009; Exley, 2016).

To account for the effect of incentives on reputational signalling, Bénabou and Tirole (2006) propose a model where an agent solves:

$$\max_{a} \quad (v_a + v_y y)a - C(a) + \mu_a E(v_a|a, y) - \mu_y E(v_y|a, y) \tag{2}$$

where v_a is the intrinsic motivation for prosocial action a, v_y the extrinsic motivation (greediness) for incentive y, $C(a)$ is the cost of the action, μ_a is the reputational concern to appear as prosocial, $E(v_a|a, y)$ is the second-order prosociality attribution [5] given the action

[4] A prosocial reputation is linked to direct evolutionary advantages such as increased mating success (Arnocky et al., 2017; Bhogal et al., 2019).

[5] In other words, second-order beliefs capture what motivation the decision maker believes that others will attribute their action to. This applies for attributions to both prosociality and greediness. In the initial example of late-coming parents to the daycare, this relates to what parents believe others would think if they pick their children up on time: are they doing the right thing or just avoiding the fine?

and incentive, μ_y is the concern to appear as not greedy, and $E(v_y|a, y)$ is the second-order belief of the attribution to agent's greediness. The existence of incentives can alter behaviour as the agent will anticipate a change in what others attribute their action to.

2.1.3 Joining both signals: a complete model

As we have seen, both signalling-in (to the agent) and out (to others) have to be considered to understand how incentives crowd out prosocial actions. To capture both mechanisms, I join the model of signalling-in by Bowles and Polania-Reyes (2012) and the model of signalling-out by Ariely et al. (2009) in a single function:

$$\max_a \quad a(\lambda_0(1 + \phi(y)\lambda_c + y\lambda_m) + v_y y) - C(a) + \mu_a E(v_a|a, y) - \mu_y E(v_y|a, y) \tag{3}$$

where all parameters are consistent with those in equations 1 and 2. In this framework, crowding-out can be produced by (1) the interaction between incentives and social preferences (λ_c, λ_m), and (2) the effect of incentives in expected attribution $(E(v_a|a, y), E(v_y|a, y))$. While the first interaction can occur both in private and in public, the second one depends on the visibility of incentives and behaviour.

2.2 Gender differences

Although gender differences in prosocial behaviour has been extensively studied, gender differences in how incentives crowd prosociality out have been largely neglected. Previous research in economics (Exley, 2016; DellaVigna et al., 2013; Mellström and Johannesson,

2008; Lacetera and Macis, 2010a) and psychology (Deci et al., 1975; Kast and Connor, 1988) provide evidence that incentives crowd out prosocial actions more for women than for men, whereas other studies don't find any significant differences (Ariely et al., 2009; Fryer, 2011; Goette and Stutzer, 2020). As there is no consensus in the literature, let's start by asking why can crowding-out differ across gender.

2.2.1 Gender differences in signalling-in

While there is no clear consensus in the literature on gender differences in prosocial behaviour (Engel, 2011; Eckel and Grossman, 1998, 2008; Niederle, 2016), there's agreement on women being "more sensitive to social cues in determining appropriate behaviour" (Croson and Gneezy, 2009, p. 16). For instance, Ben-Ner et al. (2004) find that manipulating the gender or home state of the recipient in a Dictator Game changes the amount offered by female but not male dictators. Differential sensitivity to social context in altruism is already observed in 8 years old schoolchildren. When allocating M&Ms, girls offer more to boys, while boys' offers do not depend on gender (Houser and Schunk, 2009). Previous literature thus concludes that women adapt more their prosocial behaviour to changes in the experimental setting, reference agents, priming and other contextual factors (Kahn et al., 1971; Eckel and Wilson, 2003; Croson and Gneezy, 2009; Niederle, 2016).

As incentives act as signals about the decision environment, the sensitivity to these signals can lead to different behaviours. For instance, if men are less sensitive to contextual signals, we can expect men to crowd out less in the presence of incentives. This prediction is independent of visibility, but only in a private context could this motive be disentangled from

reputation signalling. Mellström and Johannesson (2008) find that offering private payments for blood donations reduces blood supply by half among Swedish women, but it has no effect on men. Other studies report similar results where no reputational concerns are (directly) present (Lacetera and Macis, 2010a; Lazear et al., 2012; DellaVigna et al., 2013; Levitt et al., 2016; Sittenthaler and Mohnen, 2020). In conclusion, gender differences in sensitivity to signals can explain differences in crowding-out.

2.2.2 Gender differences in signalling-out

Besides differences in signalling-in, gender differences in crowding-out can also be induced by reputational signalling. In this case, the effect of gender would only be observed in public. There is some experimental evidence showing that public incentives crowd out prosociality more in women (Jones and Linardi, 2014; Jalava et al., 2015). Following the model by Bénabou and Tirole (2006), this difference can be explained through gendered reputational concerns and gendered second-order beliefs of attributions.

First, it is well established that visibility impacts prosocial behaviour more for women than for men (Engel, 2011). For example, when decisions are made visible, women tend to adhere more to the norm of the group than men (Jones and Linardi, 2014). These results go in line with the psychological literature. A meta-analysis of 133 studies using the conformity paradigm of Asch (1952) shows that women are more likely to conform to social norms and expectations when observed by others (Bond and Smith, 1996).

Second, there is significant evidence pointing to gendered second-order beliefs about prosociality. Gender stereotypes and social roles portray men as agentic and dominant,

and women as prosocial and communal (Eagly et al., 2000; Eagly, 2009; Ellemers, 2018). These stereotypes translate in gendered expectations: women are expected to behave more prosocially. Aguiar et al. (2009) separated the Dictator Game offers made by male and female dictators into two boxes labeled accordingly, and then let recipients choose the box where their allocation would be randomly picked from. Participants choose a female dictator in most cases, expecting women to behave more prosocially. How much more? Brañas-Garza et al. (2018) elicit prosociality expectations and find that female dictators are expected to offer the double than male dictators.

Differential expectations imply differential responses to prosociality by gender. As prosociality is expected from women, they are less rewarded for it, but they get punished more harshly than men if they don't comply with the prosocial norm (Ames et al., 2004; Heilman and Okimoto, 2007; Allen, 2006; Heilman, 2012). For men, the opposite happens: they are more rewarded for doing good and less punished for omission. Gendered expectations about prosociality, as well as differential rewards and punishments, develop early in childhood (Shigetomi et al., 1981) and are sustained through adolescence and adulthood (Eagly, 1987; Hall and Carter, 1999; Salgado, 2018). These processes make gendered expectations to be internalized and anticipated by individuals. For instance, Boschini et al. (2018) find women give significantly more in the Dictator Game when their gender is primed by making participants state it just before making the decision, but not in the baseline condition. This result suggests women internalize gendered expectations as second-order beliefs, in line with the gender stereotypes literature (Atkinson et al., 2021; Ellemers, 2018). This is, they believe others expect more from them. Summing up, women are expected to behave more prosocially

than men, and the gendered expectations are internalized and anticipated by individuals. In the framework of model 3, second-order beliefs are captured by gender dependent distributions of the extrinsic ($E(v_y|a, y, gender)$) and intrinsic motivations ($E(v_a|a, y, gender)$).

In conclusion, gender differences in crowding-out might arise from gendered reputational concerns and second-order beliefs. If these are the mechanisms behind gender differences, they will only be appreciable in reputation-relevant contexts, this is, when incentives are offered in public.

2.3 What we don't know: aim of this study

The reviewed research proves that incentives can crowd prosocial motivations out, reducing prosocial efforts instead of promoting them. This effect seems to be induced by signalling-in and signalling-out. Besides, it's not clear whether crowding-out varies depending on gender, and the potential mechanisms behind these gender differences have not been studied yet. To address these gaps, this study proposes three research questions: (1) Is crowding-out induced by signalling? (2) Are there gender differences in crowding-out? (3) In affirmative case, what are the mechanisms behind these gender differences? The first question replicates the findings by Ariely et al. (2009), and the following two go beyond the original paper.

3 Methods

3.1 Experimental design

The experimental design has to fulfill three conditions: the prosocial task must have image implications by allowing reputation building; incentives and participant's decisions must be made either public or private; and monetary incentives must be either offered or not.

In this experiment, participants can behave prosocially by completing a simple task that represents a donation to the Red Cross. The real-effort task consists in counting the number of ones in a 5x10 matrix of zeros and ones, a task successfully used by Abeler et al. (2011) and Exley (2016). Every matrix correctly solved implies a donation of 1 euro to the Red Cross, and participants can complete up to 15 matrices per condition (which takes around 5 minutes). Before starting, every participant practiced the task by solving two mock matrices that didn't convey any payments. Participants were informed that all payments included in this study were hypothetical.

While this task is not exactly the same used by Ariely et al. (2009), it is compatible as it has the same characteristics. First, it is a real-effort task, where participants have to work to be able to donate. Second, as the task represents a donation to a well-perceived cause [6], it allows participants to signal a prosocial reputation by doing it. Third, the task is completely superfluous, not giving the participants any motivation to do it outside the will to donate and appear as prosocial [7]. Fourth, the estimated time to solve 15 matrices is 5 minutes, the

[6] Median identification with the values and mission of the Red Cross is 7 in our sample, in a scale from 0 to 10.

[7] While there might be some participants that enjoy doing the task, this unobserved variable is uncorrelated with the treatments, which are randomly allocated.

same time participants in Ariely et al. (2009) had for the task.

Based on this task, this study uses a mixed 2x2 design. *Visibility* was manipulated between-subjects by randomly assigning participants to either a "private" or a "public" condition. In the private condition, participants were informed that their donation choice would be kept completely anonymous and private. In the public condition, they were informed that "the researchers conducting this experiment will see your name and the donations you made, analyze your choice and compare it with other respondents' choices" [8]. As the choices are indeed analyzed and compared in this study, participants were not deceived. To make the public visibility more salient, a stylized image of a pair of "watching-eyes" was included on the interface of the public condition, a manipulation shown to increase the visibility salience and prosocial activity (Bateson et al., 2006; Bradley et al., 2018; Haley and Fessler, 2005; Rigdon et al., 2009). The visibility manipulation allows to measure the effect of reputational concerns in prosocial behaviour and its interaction with incentives.

The visibility condition in this study differs from the one in the original experiment. In the latter, subjects disclosed to the other lab participants their donations, as well as how much money they had personally earned from them. This manipulation of visibility is stronger, but it was unfeasible for this study as it is conducted through an online survey. However, as Figure 3 presents, the effect of visibility is comparable across studies, suggesting that the visibility condition in this study was equally effective.

The payment scheme was manipulated within-subjects, presenting each participant with two conditions. The not-paid condition was the already described baseline, where each matrix

[8] Bénabou et al. (2020) successfully uses a similar visibility manipulation.

conveyed a one euro donation for the Red Cross. In the paid condition, each matrix solved represented one euro payment for the participant in addition to the charitable donation. The order of the paid and not-paid conditions was randomized to ensure that fatigue was independent to the treatment. Participants could solve up to 15 matrices per payment condition, making a total of 30 possible matrices. The payment scheme manipulation, which allows to measure the effect of incentives on prosocial effort, is the same as in the original study.

From this design I extract 4 different outcomes. First, the *number of paid and not paid matrices* correctly solved measure the absolute prosocial effort done in each incentive condition. Second, to easily see the effect of incentives on prosocial effort, the *ratio of correct matrices* is calculated as the number of correct paid matrices divided by the number of correct unpaid ones. A ratio of one reflects neutrality to incentives, a ratio greater than one shows a positive effect of incentives, and a ratio less than one reflects crowding-out (less prosocial action due to incentives). In the same line, the *ratio of tried matrices* includes all attempted matrices, both correct and incorrect. Finally, the *crowded out* binary variable keeps record of which participants decreased their effort in response to incentives, this is, those with a ratio of correct matrices less than one.

The 4 different outcomes tell the same story from different perspectives. As Appendix Figure 8 displays, all outcome variables are significantly correlated. All analyses and results are robust to different outcome choices. Thus, for each analysis I select the outcome that allows a better interpretation. For instance, when analyzing who crowded out and who didn't, the variable *crowded out* is most clear, and when focusing on the distribution of effectiveness

of incentives the relative outcome *ratio correct* offers the best interpretation.

After completing the practice and the two payment conditions, participants filled a short questionnaire. The questionnaire included basic demographics (gender and age) and identification with the mission and values of the Red Cross, as well as several items designed to proxy the parameters of model 3. To approximate social preferences λ_0, I used the self-reported inequity aversion item by Espín et al. (2018) based on the model by Fehr and Schmidt (1999). For reputational concerns μ_a, I employed the Others' Approval subscale of the Contingency of Self-Worth Scale (Crocker et al., 2003), with a Cronbach's alpha of 0.78. To account for second-order beliefs $E(v_a|a, y)$, participants answered two questions. The first one captured expected judgement from deviating from prosociality, using a modified version of the lost wallet paradigm (Dufwenberg and Gneezy, 2000). The second one measured the second-order belief of how much prosociality others expect the participant to show, using a modified version of the altruism measure by Falk et al. (2018). These two items have a Cronbach's alpha of 0.67. All items (shown in the Appendix) in this follow-up questionnaire were measured on a 0-10 Likert Scale, and their order within the questionnaire was randomized.

3.2 Data collection

The experiment was conducted using an online survey. I opted to run the experiment in two languages, in English and back-translated into Spanish, to be able to reach more participants and increase the number size. The survey link was shared in social media. 63 people answered to the English version and 115 to the Spanish one. Table 1 presents the

descriptive statistics grouped by survey language, as well as a Kolmogorov-Smirnov test for each variable.

Table 1: Descriptive statistics by survey language

	N	Mean	SD	N	Mean	SD	K-S: p
Language	English	English	English	Spanish	Spanish	Spanish	
Age	65	25.6	6.6	113	27.5	7.4	0.1
Female	65	0.5	0.5	113	0.5	0.5	0.9
Visibility	65	0.6	0.5	113	0.5	0.5	0.2
Not Paid	65	8.2	4.4	113	8	4.6	0.8
Paid	65	8	4.2	113	8.1	4.3	1
Ratio Correct	62	1.2	0.8	109	1.4	1.2	0.2
Ratio Tried	63	1.2	1.2	109	1.1	0.8	0.5
Crowded Out	65	0.4	0.5	113	0.4	0.5	0.7
Red Cross Identification	65	7	2.6	113	6.8	2.4	0.6
Inequity Aversion	65	6.6	2.1	113	6.5	2.1	0.7
Reputational Concern	65	5.6	1.9	113	5.3	2.2	0.4
Second Order Beliefs	65	2.8	1.9	113	2.6	1.7	0.5

The decision to run the experiment in two languages creates a trade-off between increasing the sample size and drawing from different populations. A first question to address is whether the subject pools that answered to each version are comparable in terms of demographics and responses. The tests of equality of distributions yield no significant differences between languages. Besides, the visibility manipulation had the same behavioral effects, as shown in Appendix Figure 9, and the experimental results are similar across languages (Figure 3). The participants in each survey are comparable in terms of demographics, responses

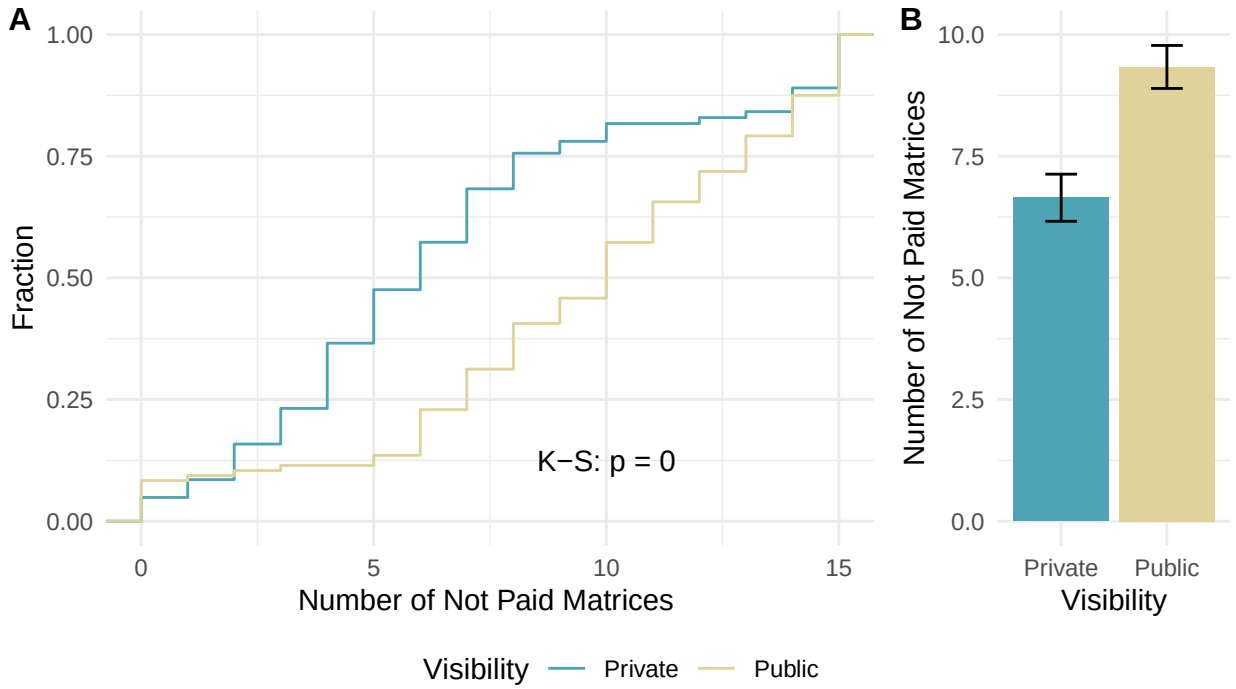
and outcomes, so increasing the sample size didn't compromise the homogeneity of the subject pool. Therefore, we consider both subject pools comparable and analyze them jointly henceforth, making a total of 168 observations.

A second question is whether the two experimental manipulations, *visibility* and *incentives*, had the intended effect on prosocial effort. Figure 1 shows the positive effect of visibility on the number of not paid matrices, to separate it from its interaction with incentives. Visibility enhances prosocial behaviour when no monetary extrinsic motivation is present, in line with the hypothesis that prosocial activity functions as a reputation signal. Table 3 confirms the main effect of visibility (*Public*) is positive and significant across models, using OLS regression. Thus, we can conclude that the visibility manipulation had the intended effect, increasing prosocial activity when no incentives are offered, in line with Ariely et al. (2009).

Regarding the incentives manipulation, Table 3 shows that offering incentives in private (*Incentives*) resulted in 1.8 more solved matrices (an increase of 0.4 SD). Even though payments were hypothetical, participants behaved *as if* they were real. This way, the incentives manipulation had the intended positive effect on prosocial effort when offered in private. Also, we can interpret this result as evidence against a strong categorical crowding-out by signalling-in: offering incentives when "nobody is watching" does not decrease prosocial activity on average [9]. If there is any crowding-out by signalling-in, it is not strong enough to make incentives generally counter-effective.

[9]However, 20% of participants do decrease their prosocial effort when offered private incentives.

Figure 1: Effect of visibility on prosocial behavior. Error bars: SE of the mean.



Notes: In Panel B, error bars are the standard errors of the mean.

4 Results

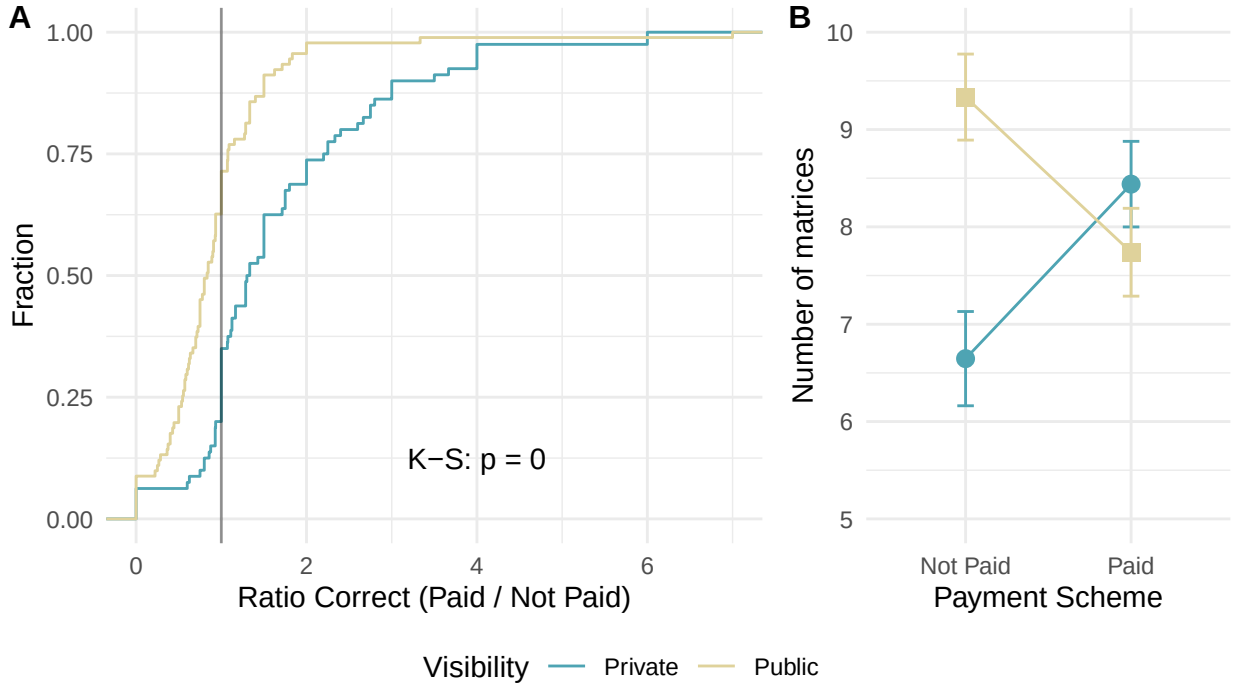
This section presents the results of the study separately for each of the research questions.

4.1 Crowding-out

Our first research question is whether crowding-out is induced by signalling-out. If this is the case, incentives offered in public will have less effectiveness than in private and cause more crowding-out. Figure 2 evidences this is indeed the case: public incentives are less effective than private ones. Panel A displays the cumulative distribution functions (CDFs) of the ratio

of correct matrices (the number of paid correct matrices divided by the number of unpaid correct) by visibility condition. This ratio shows the effect of incentives on prosocial behaviour, where a ratio less than 1 (value represented by the vertical grey line) reflects crowding-out. Panel A shows that public visibility diminishes notably the effect of incentives. In the private condition, 20% of the participants crowd out, i.e., they do less effort when it is incentivised than when it isn't. In public, 63% of participants crowd out. A Kolmogorov-Smirnov test of equality of distributions confirms that the effect of incentives varies significantly with visibility ($p = 0$). Panel B of Figure 2 presents the interaction effect of incentives and visibility in a similar way to that shown in Ariely et al. (2009). On average, incentives enhance prosocial effort in the private condition, whereas they have a negative effect in the public condition. These results confirm that crowding-out from prosocial behaviour is induced by reputation signalling.

Figure 2: Effect of incentives on prosocial behavior.



Notes: In Panel B, error bars are the standard errors of the mean.

Table 3 presents the results of 4 OLS regressions confirming the trends shown in Figure 2: public visibility diminishes the effectiveness of incentives. Model 1 only includes the experimental treatments and their interaction, Model 2 adds the demographics as controls, Model 3 replicates the specification in Ariely et al. (2009), and Model 4 includes all covariates. All models include a control for the survey language.

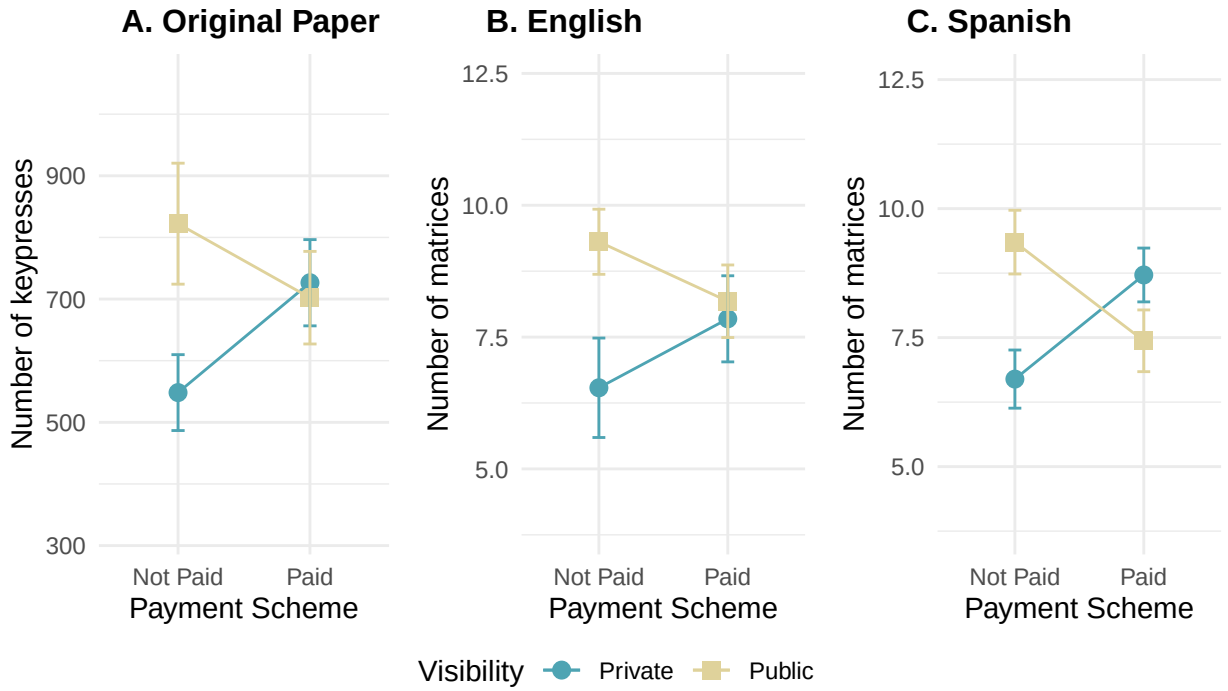
Table 3: Crowding-out induced by reputation signalling.

	Number of Matrices			
	(1)	(2)	(3)	(4)
Public	2.689***	2.714***	2.642***	2.324***
	(0.647)	(0.641)	(0.639)	(0.636)
Incentives	1.793***	1.793***	1.793***	1.793***
	(0.671)	(0.665)	(0.661)	(0.649)
Public * Incentives	−3.386***	−3.386***	−3.386***	−3.386***
	(0.914)	(0.905)	(0.900)	(0.884)
Female		1.026**	1.100**	0.424
		(0.452)	(0.451)	(0.475)
Age		−0.059*	−0.067**	−0.079**
		(0.032)	(0.032)	(0.032)
Red Cross			0.204**	0.143
			(0.092)	(0.094)
Reputn. Concern				−0.108
				(0.106)
Second Order Beliefs				0.520***
				(0.137)
Ineq. Aversion				0.109
				(0.109)
Language	0.025	0.150	0.200	0.252
	(0.475)	(0.474)	(0.472)	(0.465)
Constant	6.629***	7.597***	6.365***	6.084***
	(0.575)	(1.032)	(1.169)	(1.401)
Observations	178	178	178	178
Adjusted R^2	0.039	0.058	0.068	0.100
Note:				*p<0.1; **p<0.05; ***p<0.01

Across all specifications, public visibility (*Public*) has a positive and significant effect on prosocial effort when this is not paid. Private incentives (*Incentives*) also have a positive effect, but of less magnitude: an incentive of 1 euro results in 1.8 ($SE = 0.7$) more matrices solved. Most importantly, visibility and incentives have a negative interaction across models.

Public visibility inverses the effect of incentives, which hinder prosocial effort ($\beta = -1.6$, $SE = 0.6$). [10] Regarding the rest of the covariates, female participants donate more across conditions. However, this difference becomes non-significant when second-order prosociality beliefs are accounted for, having the latter a positive relation with donations. We will further explore this issue in Section 4.3. The rest of the covariates included don't show any significant relation with the number of matrices solved.

Figure 3: Crowding-out in the original paper, and in this study by survey language



Notes: To avoid misleading visualization, the range of the y-axis of each of the 3 panels is set as the panel mean plus or minus its standard deviation. Error bars are standard errors of the mean.

Figure 3 demonstrates the main results of this experiment are parallel to those of the original study, and also comparable across survey languages. Using the data from the original study, openly available at the AER website, this figure shows that the effectiveness

[10] Appendix Table 5 shows the negative impact of visibility on the effectiveness of incentives holds even if we consider relative outcomes (ratios of correct and attempted matrices) instead of the absolute number of matrices solved.

of incentives on prosocial effort decreases with visibility in a similar manner in the original study and in the two languages of this experiment. This result replicates the main finding of Ariely et al. (2009).

To sum up, visibility has a negative and significant impact on the effectiveness of incentives on prosocial behaviour. Crowding-out of prosociality is of greater magnitude in public than in private, evidencing that crowding-out is induced by reputation signalling-out. These results replicate the original study by Ariely et al. (2009) and contribute to the crowding-out literature.

4.2 Gender differences

Going beyond the replication of Ariely et al. (2009), the second research question of this study adresses whether there exist gender differences in crowding-out. Section 2.2 presents evidence from previous research supporting the hypothesis that women crowd out more of prosocial activities, both in private (induced by signalling-in) and public (induced by signalling-out). However, no research focuses on establishing these differences. In this section, we will focus on them and show that women crowd out more of prosocial behaviours in the presence of public incentives.

Figure 4 presents the CDFs of the ratio of correct matrices separated by visibility condition and gender, showing there are gender differences in public. In the private condition (Panel A), incentives crowd prosociality out for 19% of men and 21% of women. A Kolmogorov-Smirnov test yields no distributional differences across gender. In the public condition, displayed in Panel B, gender differences can be appreciated. While 48% of male participants crowd out

in public, 77% of female participants do so. The difference in distributions is significant at the 5% level. This way, Figure 4 provides evidence of gender differences in crowding-out, as public visibility has a higher impact on the effectiveness of incentives for women than for men. As the difference is only observed in public, it seems that it is induced by signalling-out mechanisms.

Figure 4: Gendered crowding-out.

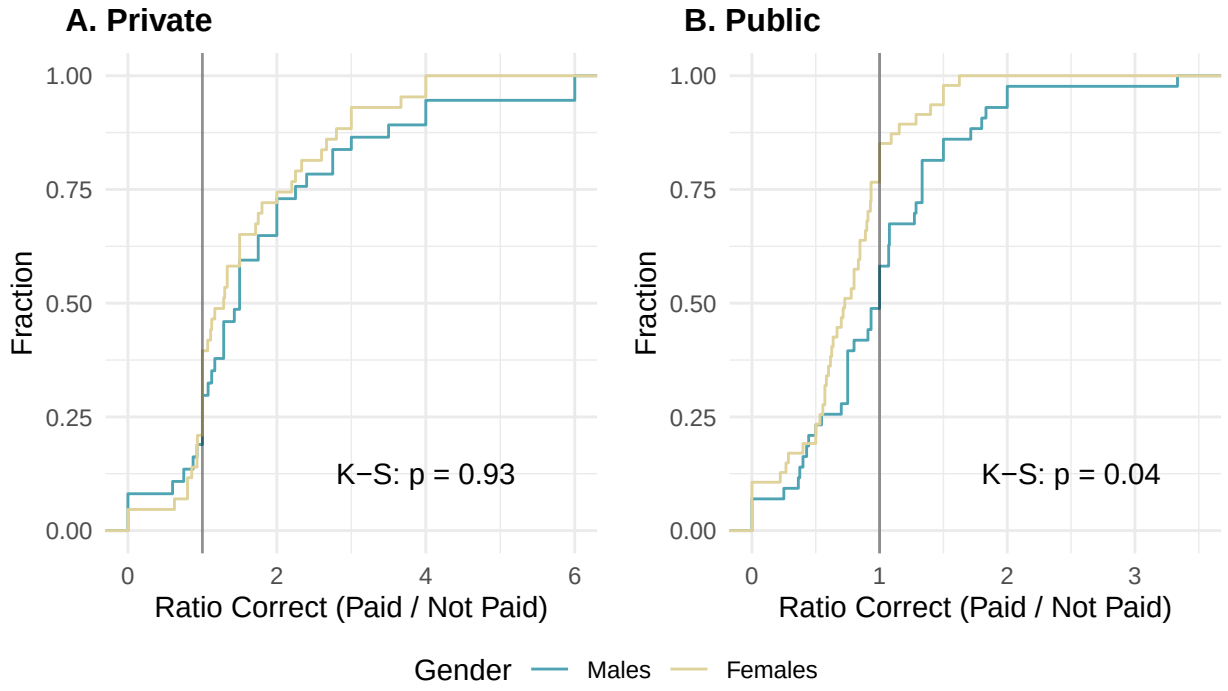


Table 4 validates, across 5 OLS models, that public incentives crowd prosociality out more for females than for males. Model 1 and Model 2 divide the complete specification between male and female participants, respectively. Models 3 and 4 use the ratio of paid/unpaid correct matrices as dependent variable, the latter including all covariates. Model 5 uses as dependent variable a dummy that equals 1 if the participant crowded out (ratio less than 1). Models 1 and 2 show that the crowding-out coefficient (*Public* * *Incentives*) is more than

double for women than for men. [11]. This difference is also observable in Models 3 and 4, as the interaction *Public* * *Female* is negative, though not significant in this case.

Table 4: Gender differences in crowding-out induced by signalling.

	Number of Matrices		Ratio Correct		Crowded Out
	Males	Females			
	(1)	(2)	(3)	(4)	(5)
Public	2.256**	2.633***	−0.660**	−0.602**	0.211**
	(0.973)	(0.874)	(0.288)	(0.285)	(0.097)
Incentives	1.842**	1.750**			
	(0.922)	(0.783)			
Public * Incentives	−2.217*	−4.562***			
	(1.288)	(1.186)			
Female			−0.240	−0.112	−0.053
			(0.269)	(0.276)	(0.091)
Public * Female			−0.205	−0.064	0.270**
			(0.328)	(0.332)	(0.136)
Age	−0.039	−0.117***	−0.018*	−0.019*	0.006
	(0.058)	(0.045)	(0.010)	(0.010)	(0.005)
Red Cross	0.153	0.115		−0.007	−0.003
	(0.109)	(0.133)		(0.028)	(0.013)
Reputn. Concern	−0.060	−0.147		−0.004	0.001
	(0.174)	(0.137)		(0.050)	(0.016)
Second Order Beliefs	0.358	0.667***		−0.173***	0.075***
	(0.274)	(0.192)		(0.044)	(0.020)
Ineq. aversion	−0.091	0.227*		0.046	−0.006
	(0.168)	(0.134)		(0.032)	(0.016)
Language	−0.547	1.233*	0.143	0.118	−0.002
	(0.700)	(0.637)	(0.157)	(0.151)	(0.069)
Constant	6.466***	6.184***	2.222***	2.357***	−0.057
	(2.212)	(2.175)	(0.318)	(0.425)	(0.202)
Observations	86	92	171	171	178
Adjusted R^2	0.030	0.158	0.144	0.205	0.249

Note: *p<0.1; **p<0.05; ***p<0.01
HC-Robust SE in parentheses

In Model 5 the interaction of public visibility and gender is positive and significant. That is, while public visibility increases by 21 points ($SE = 10$) the percentage of men that crowd out, for women this effect is more than double (48 points, $SE = 10$). While there are

[11]The difference between coefficients ($\Delta\beta = -2.3, SE = 0.6$) is significant at the 1% level ($t = -3.7$) The calculation for the standard error follows the method by Clogg et al. (1995). However, it must be noted that this method was developed for nested models, and in this case the models are non-nested. The comparison is thus only orientative.

some differences in statistical significance across models, all specifications point to the same direction: public incentives crowd prosociality out more for women than for men.

In conclusion, these results validate the existence of gender differences in crowding-out. Specifically, public incentives crowd women out more of prosocial actions than men, while this effect is not appreciable for private incentives. This contrast depending on visibility points to gender differences being caused by signalling-out and not signalling-in. Next section further explores the mechanisms behind gender differences.

4.3 Mechanisms

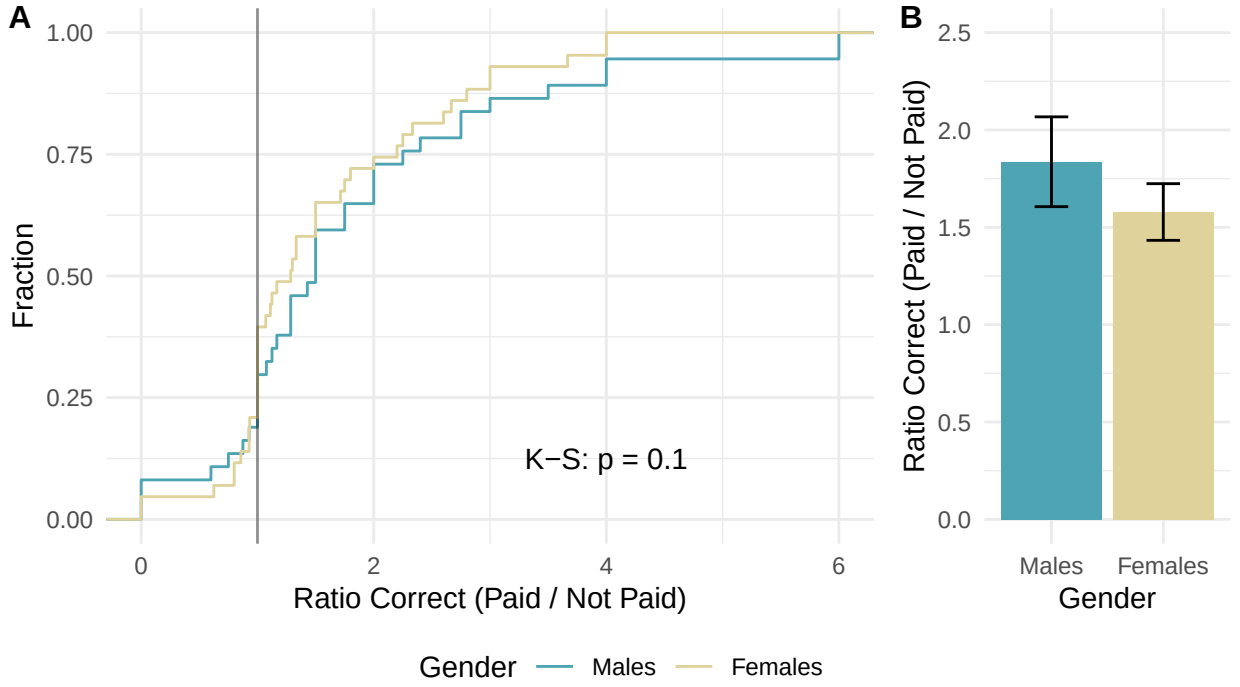
The previous section proves that crowding-out induced by signalling is of greater magnitude for women than for men. This section addresses the mechanisms behind these gender differences.

Previous research, aggregated into Model 3, posits two different mechanisms for crowding-out: signalling-in and signalling-out. Signalling-in considers that incentives act as signals about the decision environment for the agent. This mechanism thus depends on the sensitiveness to contextual signals during decision making, which is usually more pronounced in women (see Section 2.2.1). If this is the reason for gender differences in crowding-out, we would expect the effectiveness of private incentives, when no reputational concerns are directly present, to vary across genders. Specifically, if incentives signal a change from a social to a market environment, and women adapt more to these signals, private incentives will be less effective for women.

Figure 5 demonstrates this is not the case in our data, as the effect of incentives in the

private condition is similar for male and female participants. Panel A shows the CDFs of the ratio of correct matrices. No distributional differences can be appreciated between males and females. Panel B presents the first moment of these distributions. Again, there is no significant difference between genders. Table 4 offers more evidence against gender differences being induced by signalling-in. There is no observable difference between the effect of private incentives in models 1 and 2 (*Incentives*), nor an effect of gender (*Female*) over incentive effectiveness (models 3 and 4) or crowding-out probability (model 5). In conclusion, there is no support for the hypothesis that gender differences in crowding-out arise from differences in sensitiveness to signals.

Figure 5: Signalling-in by gender (private behaviour).



Notes: In Panel B, error bars are standard errors of the mean.

Observed differences in crowding-out can also be caused by differences in signalling-out. As discussed in Section 2.2.2, behaving prosocially acts as a reputation signal. The presence of

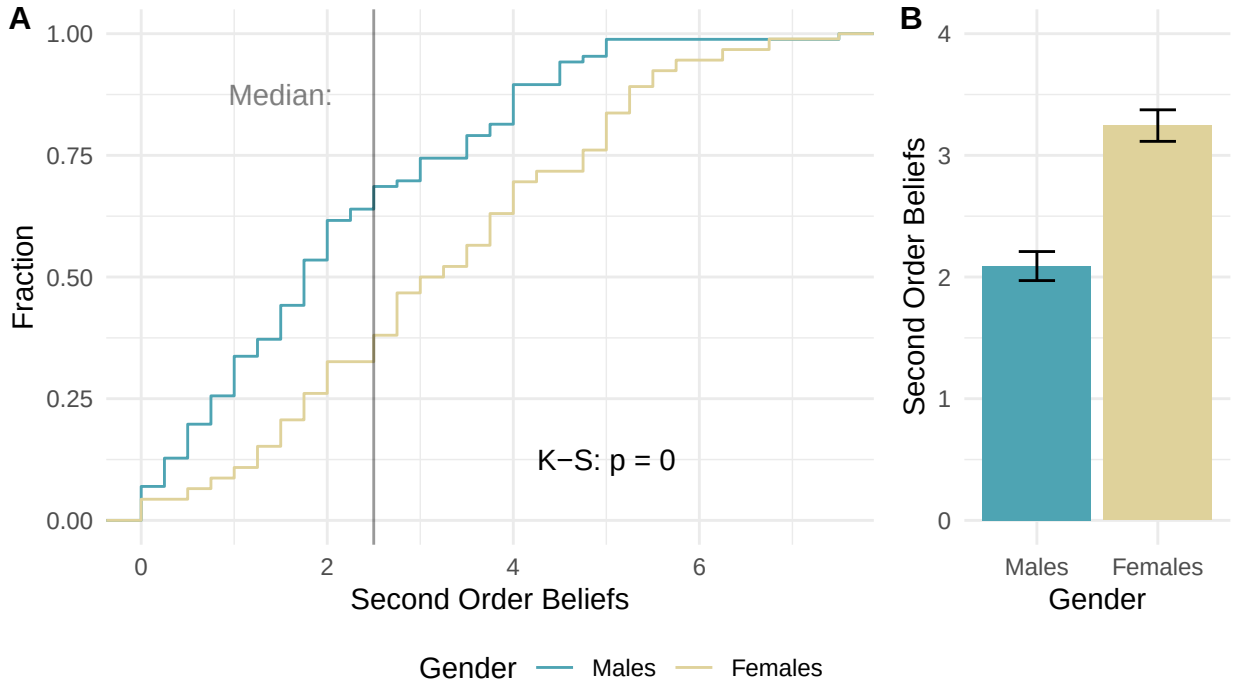
incentives increases the noise in this signal and can even change its content, as the agent could be behaving prosocially not out of good intentions but out of greediness. Following Model 3, signalling-out is determined by the interaction of two parameters: reputational concern (μ_a) and the second-order belief of what the action would be attributed to ($E(v_a|a, y)$).

Are gender differences in crowding-out caused by differences in reputational concern? Using the Others' Approval subscale as a measure of self-reported reputational concern, there are no observable gender differences in means (t-test, $p = 0.8$) or distributions (K-S test, $p = 1$). It must be noted that this scale measures general reputational concern and does not focus solely on the desire to appear as prosocial. However, as prosociality is the social norm, we expect an individual concerned with their reputation in general to be concerned about appearing as prosocial. In any case, reputational concern doesn't explain the observed differences in the data.

While there seems to be no differences in reputational concerns, previous research agrees that the same behaviour is attributed to different motivations depending on gender (see Section 2.2.2). Individuals anticipate these attributions and therefore believe that others expect more or less prosociality from them based on their gender. Figure 6 shows this pattern is also present in our sample: women believe more prosociality is expected from them. Panel A shows this difference in second-order beliefs exists all along the distribution, and it is also reflected in the difference in means by gender in Panel B. While 62% of women hold beliefs above the sample median, only 31% of men do so. Women believe more is expected from them in terms of prosocial behaviour, so, in order to signal a 'good' (better than expected) reputation, they feel they have to put more effort than men.

Do gender differences in second-order prosociality beliefs explain heterogeneity in crowding-out? If this is the case, we would expect the effect of gender to dilute when second-order beliefs are accounted for. The evidence displayed in Table 4 supports this prediction. Controlling for second-order beliefs in the regression model reduces the point estimate of the effect of gender on incentive effectivenes in public from -0.2 ($SE = 0.3$) to -0.06 ($SE = 0.3$) (Models 3 and 4, respectively). This points to an overestimation of the effect of gender due to omitted variable bias, as beliefs are correlated both with gender and the outcome variable. However, Model 5 shows there are still gender differences in crowding-out when controlling for second-order beliefs. Overall, although second-order prosociality beliefs might not be the whole story, they explain gender differences in crowding-out.

Figure 6: Gender and second-order prosociality beliefs.

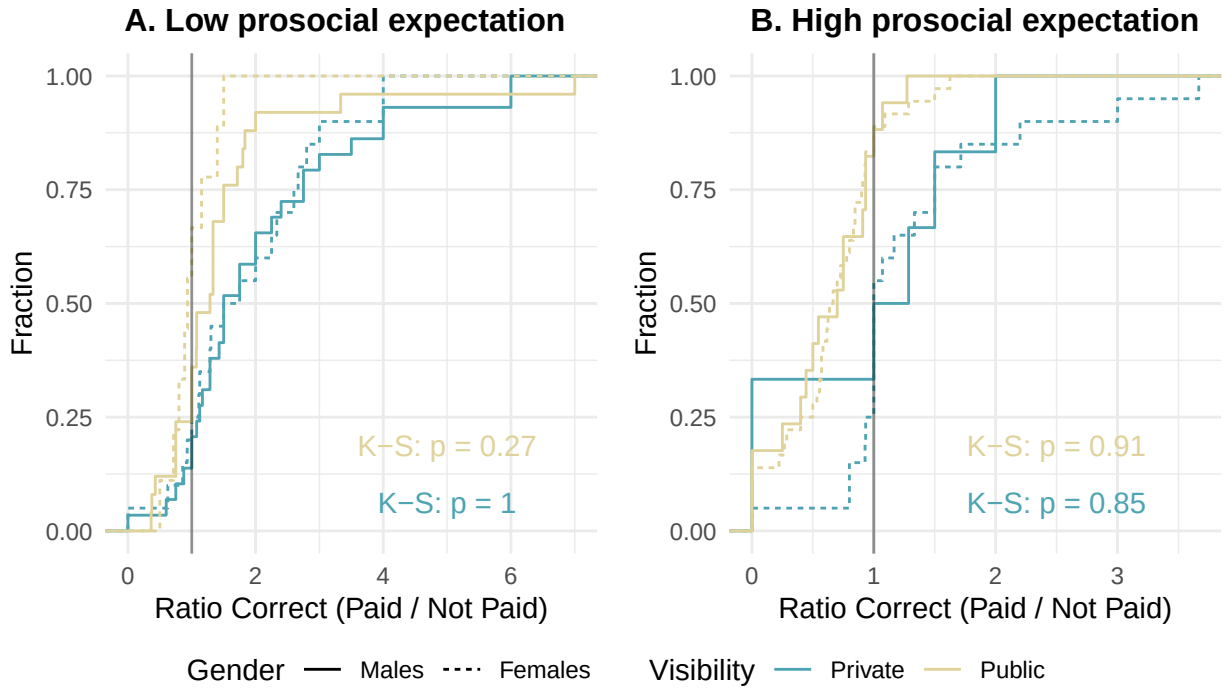


Notes: In Panel B, error bars are standard errors of the mean.

Figure 7 presents more evidence on second-order prosociality beliefs or expectations

as the key mechanism behind gender differences in crowding-out. Panel A displays the CDFs for the ratio of correct matrices for those participants at or below the median of second-order beliefs. While 63% of all participants crowd out in public, among those with low second-order expectations only 34% crowd out. Two Kolmogorov-Smirnov tests show there are no significant differences by gender in responses to incentives for those with low second-order prosocial expectations. In the same line, Panel B shows that gender differences in crowding-out disappear for those in the high prosocial expectation group. Men and women in this group react the same way to incentives in public (K-S test, $p = 0.9$), with around 83% of them reducing their prosocial effort when offered public incentives. These results illustrate that second-order prosociality beliefs explain gender differences in crowding-out of prosocial behaviour.

Figure 7: Crowding-out, by second-order beliefs and gender.



5 Conclusions

Although previous research has studied how extrinsic incentives crowd prosociality out, some questions remain answered. This experimental study closes the gap by addressing three of of these questions. First, the results of the experiment show that crowding-out is induced by reputation signalling-out. Public visibility reduces the effectiveness of incentives to enhance prosocial effort to the point of making them countereffective. This result replicates the main finding of the study by Ariely et al. (2009). Second, there exist gender differences in crowding-out. Public incentives crowd prosocial effort out more for women than for men. Third, this difference is induced by gendered reputation signalling. Women believe that, everything being equal, they will be judged more harshly for deviating from prosociality than men, and that to be considered prosocial, they have to put more effort. The difference in second-order beliefs of others' attributions explains the gendered response to incentives for prosocial behaviour.

These results have important policy implications. For instance, if a charity considers incentivizing donations with monetary vouchers, it should expect it to work better if donations are kept private and not publicized. Also, offering public incentives for blood donation can be expected to be less effective than private rewards after the donation. Knowing that monetary incentives crowd out prosocial effort, symbolic instead of monetary rewards, such as public recognition, can be more effective in public settings as they would contribute to social image formation instead of mining it (Heyman and Ariely, 2004; Lacetera and Macis, 2010b).

This study was designed to experimentally test the existence of crowding-out induced by signalling-out, so the findings beyond that are correlational and not experimental. Especially,

the study of the mechanisms behind gender differences relies on correlations and distributions conditional on self-reported second-order beliefs. This method does not allow to infer causality. To address this limitation, a future study experimentally manipulating second-order beliefs would be welcome to ascertain if second-order beliefs *cause* and not only *explain* gender differences. Also, additional research on how image motivation is constructed and how this process varies across gender would shed some light onto this topic.

6 References

Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference Points and Effort Provision. *American Economic Review*, 101(2):470–92.

Aguiar, F., Brañas-Garza, P., Cobo-Reyes, R., Jimenez, N., and Miller, L. M. (2009). Are women expected to be more generous? *Experimental Economics*, 12(1):93–98.

Allen, T. D. (2006). Rewarding Good Citizens: The Relationship Between Citizenship Behavior, Gender, and Organizational Rewards1. *Journal of Applied Social Psychology*, 36(1):120–143.

Ames, D. R., Flynn, F. J., and Weber, E. U. (2004). It's the Thought That Counts: On Perceiving How Helpers Decide to Lend a Hand:. *Personality and Social Psychology Bulletin*, 30(4):461–474.

Andreoni, J. (1990). Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economic Journal*, 100(401):464.

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

Ariely, D., Bracha, A., and Meier, S. (2009). Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. *American Economic Review*, 99(1):544–55.

Arnocky, S., Piché, T., Albert, G., Ouellette, D., and Barclay, P. (2017). Altruism predicts mating success in humans. *British Journal of Psychology*, 108(2):416–435.

Arrow, K. J. (1972). Gifts and Exchanges. *Philosophy and Public Affairs*, 1(4):343–362.

Asch, S. (1952). *Social Psychology*. Prentice-Hall, Englewood Cliffs, NJ.

Atkinson, C., Buie, H., Sandstrom, G., Aknin, L., and Croft, A. (2021). Testing the GRIP: An Empirical Examination of the Gender Roles Inhibiting Prosociality Model. *Sex Roles*, 85(7-8):440–462.

Bandura, A. (1991). Social Cognitive Theory of Moral Thought and Action. In Kurtines, W. M. and Gewirtz, J., editors, *Handbook of Moral Behavior and Development*. Psychology Press.

Bateson, M., Nettle, D., and Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, 2(3):412–414.

Ben-Ner, A., Kong, F., and Putterman, L. (2004). Share and share alike? Gender-pairing, personality, and cognitive ability as determinants of giving. *Journal of Economic Psychology*, 25(5):581–589.

Bénabou, R., Falk, A., Henkel, L., and Tirole, J. (2020). Eliciting Moral Preferences: Theory and Experiment.

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

Bhogal, M. S., Farrelly, D., and Galbraith, N. (2019). The role of prosocial behaviors in mate choice: A critical review of the literature. *Current Psychology*, 38(4):1062–1075.

Bond, R. and Smith, P. B. (1996). Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychological Bulletin*, 119(1):111–137.

Boschini, A., Dreber, A., von Essen, E., Muren, A., and Ranehill, E. (2018). Gender and altruism in a random sample. *Journal of Behavioral and Experimental Economics*, 77:72–77.

Bowles, S. and Polania-Reyes, S. (2012). Economic Incentives and Social Preferences: Substitutes or Complements? *Journal of Economic Literature*, 50(2):368–425.

Bradley, A., Lawrence, C., and Ferguson, E. (2018). Does observability affect prosociality? *Proceedings of the Royal Society B: Biological Sciences*, 285(1875).

Brañas-Garza, P., Capraro, V., and Rascón-Ramírez, E. (2018). Gender differences in altruism on Mechanical Turk: Expectations and actual behaviour. *Economics Letters*, 170:19–23.

Clogg, C. C., Petkova, E., and Haritou, A. (1995). Statistical Methods for Comparing Regression Coefficients Between Models. *American Journal of Sociology*, 100(5):1261–1293.

Crocker, J., Luhtanen, R. K., Lynne Cooper, M., and Bouvrette, A. (2003). Contingencies of Self-Worth in College Students: Theory and Measurement. *Journal of Personality and Social Psychology*, 85(5):894–908.

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

Deci, E., Koestner, R., and Ryan, R. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin*, 125(6):627–668.

Deci, E. L., Cascio, W. F., and Krusell, J. (1975). Cognitive evaluation theory and some comments on the Calder and Staw critique. *Journal of Personality and Social Psychology*, 31(1):81–85.

DellaVigna, S., List, J. A., Malmendier, U., and Rao, G. (2013). The Importance of Being Marginal: Gender Differences in Generosity. *American Economic Review*, 103(3):586–90.

Dufwenberg, M. and Gneezy, U. (2000). Measuring Beliefs in an Experimental Lost Wallet Game. *Games and Economic Behavior*, 30(2):163–182.

Eagly, A. (1987). *Sex Differences in Social Behavior: A Social-role interpretation*. Erlbaum, London.

Eagly, A., Wood, W., and Diekman, A. (2000). Social role theory of sex differences and similarities: A current appraisal . In Eckes, T. and Trautner, H., editors, *The developmental social psychology of gender*. Erlbaum, New Jersey.

Eagly, A. H. (2009). The His and Hers of Prosocial Behavior: An Examination of the Social Psychology of Gender. *American Psychologist*, 64(8):644–658.

Eckel, C. and Wilson, R. (2003). Conditional Trust: Sex, Race, and Facial Expressions in a Trust Game. In Ostrom, E. and Walker, J., editors, *Trust and Reciprocity: Interdisciplinary lessons from experimental research*. Russell Sage Foundation, New York.

Eckel, C. C. and Grossman, P. J. (1998). Are Women Less Selfish Than Men?: Evidence From Dictator Experiments. *The Economic Journal*, 108(448):726–735.

Eckel, C. C. and Grossman, P. J. (2008). Differences in the Economic Decisions of Men and

Women: Experimental Evidence. In Plott, C. R. and Smith, V. L., editors, *Handbook of Experimental Economics Results*, volume 1, pages 509–519. Elsevier.

Ellemers, N. (2018). Gender Stereotypes. *Annual Review of Psychology*, 69:275–298.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14(4):583–610.

Espín, A., Moreno-Herrero, D., Sánchez-Campillo, J., and Rodríguez Martín, J. A. (2018). Do envy and compassion pave the way to unhappiness? Social preferences and life satisfaction in a Spanish city. *Journal of Happiness Studies*, 19(2):443–469.

Exley, C. (2016). Incentives for Prosocial Behavior: The Role of Reputations. *Management Science*, 64(5):2460–2471.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global Evidence on Economic Preferences. *The Quarterly Journal of Economics*, 133(4):1645–1692.

Fehr, E. and Falk, A. (2002). Psychological foundations of incentives. *European Economic Review*, 46(4-5):687–724.

Fehr, E. and Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, 14(3):159–181.

Fehr, E. and Krajbich, I. (2013). Social Preferences and the Brain. In Glimcher, P. W. and Fehr, E., editors, *Neuroeconomics: Decision Making and the Brain: Second Edition*, chapter 11, pages 193–218. Elsevier Inc., 2 edition.

Fehr, E. and Rockenbach, B. (2003). Detrimental effects of sanctions on human altruism. *Nature*, 422(6928):137–140.

Fehr, E. and Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.

Fryer, R. G. (2011). Financial Incentives and Student Achievement: Evidence from Randomized Trials. *The Quarterly Journal of Economics*, 126(4):1755–1798.

Gneezy, U., Meier, S., and Rey-Biel, P. (2011). When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives*, 25(4):191–210.

Gneezy, U. and Rustichini, A. (2000). A Fine is a Price. *The Journal of Legal Studies*, 29(1).

Goette, L. and Stutzer, A. (2020). Blood donations and incentives: Evidence from a field experiment. *Journal of Economic Behavior and Organization*, 170:52–74.

Haley, K. J. and Fessler, D. M. (2005). Nobody's watching?: Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, 26(3):245–256.

Hall, J. A. and Carter, J. D. (1999). Gender-stereotype accuracy as an individual difference. *Journal of Personality and Social Psychology*, 77(2):350–359.

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior*, 32:113–135.

Heilman, M. E. and Okimoto, T. G. (2007). Why are women penalized for success at male tasks?: The implied communality deficit. *Journal of Applied Psychology*, 92(1):81–92.

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., and

Ziker, J. (2010). Markets, Religion, Community Size, and the Evolution of Fairness and Punishment. *Science*, 327(5972):1480–1484.

Heyman, J. and Ariely, D. (2004). Effort for Payment: A Tale of Two Markets. *Psychological Science*, 15(11):787–793.

Holmås, T. H., Kjerstad, E., Lurås, H., and Straume, O. R. (2010). Does monetary punishment crowd out pro-social motivation? A natural experiment on hospital length of stay. *Journal of Economic Behavior and Organization*, 75(2):261–267.

Houser, D. and Schunk, D. (2009). Fairness, competition and gender: Evidence from German schoolchildren. *Journal of Economic Psychology*, 30(4):634–641.

Jalava, N., Joensen, J. S., and Pellas, E. (2015). Grades and rank: Impacts of non-financial incentives on test performance. *Journal of Economic Behavior and Organization*, 115:161–196.

Jones, D. and Linardi, S. (2014). Wallflowers: Experimental evidence of an aversion to standing out. *Management Science*, 60(7):1757–1771.

Kahn, A., Hottes, J., and Davis, W. L. (1971). Cooperation and optimal responding in the Prisoner's Dilemma game: Effects of sex and physical attractiveness. *Journal of Personality and Social Psychology*, 17(3):267–279.

Kamenica, E. (2012). Behavioral Economics and Psychology of Incentives. *Annual Review of Economics*, 4:427–452.

Kast, A. and Connor, K. (1988). Sex and Age Differences in Response to Informational and Controlling Feedback:. *Personality and Social Psychology Bulletin*, 14(3):514–523.

Kurzban, R., Burton-Chellew, M. N., and West, S. A. (2015). The Evolution of Altruism in Humans. *Annual Review of Psychology*, 66:575–599.

Lacetera, N. and Macis, M. (2010a). Do all material incentives for pro-social activities backfire? The response to cash and non-cash incentives for blood donations. *Journal of Economic Psychology*, 31(4):738–748.

Lacetera, N. and Macis, M. (2010b). Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme. *Journal of Economic Behavior and Organization*, 76(2):225–237.

Lazear, E. P., Malmendier, U., and Weber, R. A. (2012). Sorting in experiments with application to social preferences. *American Economic Journal: Applied Economics*, 4(1):136–163.

Lepper, M. R., Sagotsky, G., Dafoe, J. L., and Greene, D. (1982). Consequences of superfluous social constraints: Effects on young children's social inferences and subsequent intrinsic interest. *Journal of Personality and Social Psychology*, 42(1):51–65.

Levitt, S. D., List, J. A., Neckermann, S., and Sadoff, S. (2016). The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance. *American Economic Journal: Economic Policy*, 8(4):183–219.

Li, J., Xiao, E., Houser, D., and Montague, P. R. (2009). Neural responses to sanction

threats in two-party economic exchange. *Proceedings of the National Academy of Sciences*, 106(39):16835–16840.

McLean, I. and Poulton, J. (1986). Good Blood, Bad Blood, and the Market: The Gift Relationship Revisited. *Journal of Public Policy*, 6(4):431–445.

Mellström, C. and Johannesson, M. (2008). Crowding Out in Blood Donation: Was Titmuss Right? *Journal of the European Economic Association*, 6(4):845–863.

Niederle, M. (2016). Gender. In Kagel, J. H. and Roth, A. E., editors, *The handbook of experimental economics. Volume 2*, pages 481–547. Princeton University Press.

Rigdon, M., Ishii, K., Watabe, M., and Kitayama, S. (2009). Minimal social cues in the dictator game. *Journal of Economic Psychology*, 30(3):358–367.

Roberts, G., Raihani, N., Bshary, R., Manrique, H. M., Farina, A., Samu, F., and Barclay, P. (2021). The benefits of being seen to help others: indirect reciprocity and reputation-based partner choice. *Philosophical Transactions of the Royal Society B*, 376(1838).

Ross, L. and Nisbett, R. (2011). *The Person and the Situation: Perspectives of Social Psychology*. Pinter and Martin, London, 2 edition.

Salgado, M. (2018). Gender-Biased Expectations of Altruism in Adolescents. *Frontiers in Psychology*, 0(APR):484.

Shigetomi, C. C., Hartmann, D. P., and Gelfand, D. M. (1981). Sex differences in children's altruistic behavior and reputations for helpfulness. *Developmental Psychology*, 17(4):434–437.

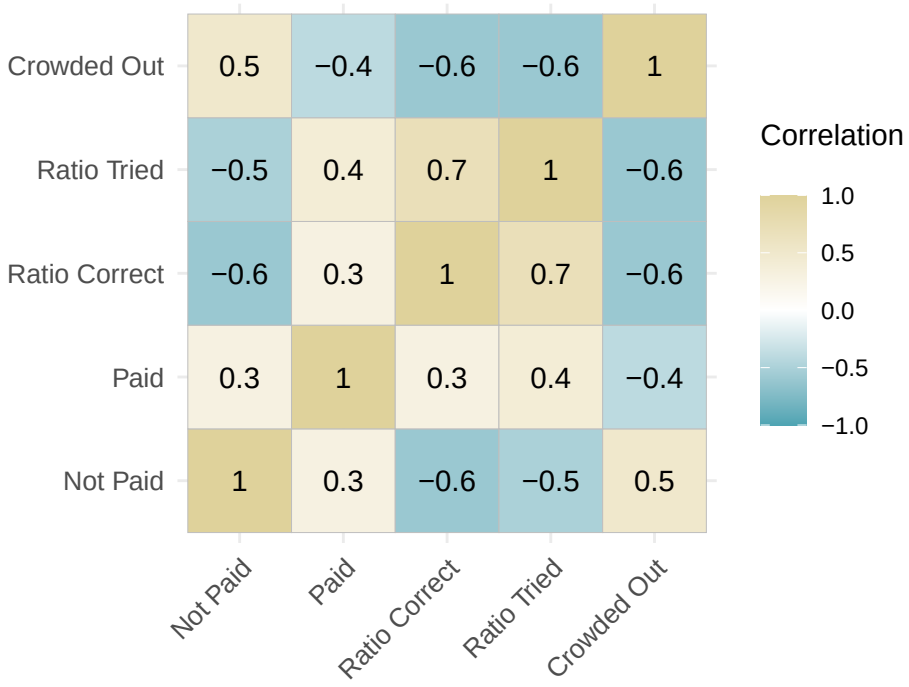
Sittenthaler, H. M. and Mohnen, A. (2020). Cash, non-cash, or mix? Gender matters! The impact of monetary, non-monetary, and mixed incentives on performance. *Journal of Business Economics 2020 90:8*, 90(8):1253–1284.

Titmuss, R. (1970). *The Gift Relationship: From Human Blood to Social Policy*. George Allen and Unwin, London.

Tversky, A. and Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, 211(4481):453–458.

7 Appendix

Figure 8: Correlation between outcomes



Notes: All correlations are significant at the 1% level.

Figure 9: Language comparison

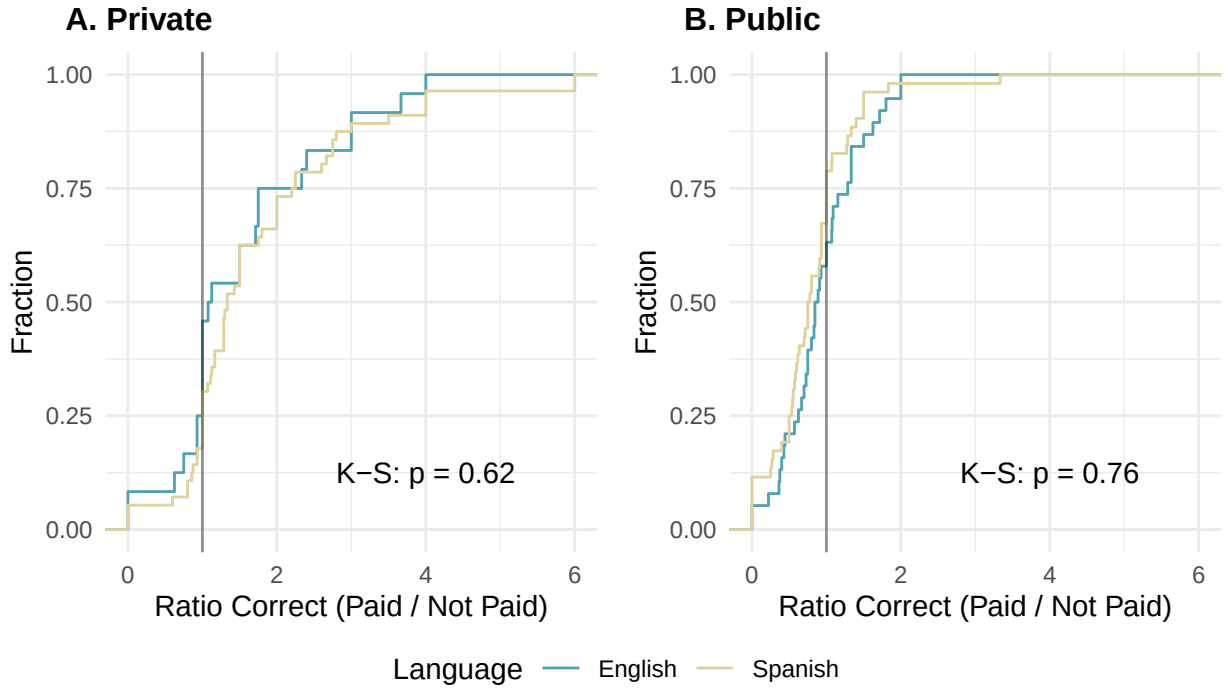


Table 5: Crowding out, relative outcomes

	Ratio Correct		Ratio Tried	
	(1)	(2)	(3)	(4)
Public	-0.767^{***}	-0.635^{***}	-0.746^{***}	-0.656^{***}
	(0.161)	(0.154)	(0.144)	(0.133)
Female	-0.352^{**}	-0.146	-0.181	-0.064
	(0.159)	(0.141)	(0.131)	(0.135)
Age	-0.018^{*}	-0.019^{*}	0.001	-0.002
	(0.010)	(0.010)	(0.009)	(0.009)
Red Cross	-0.009	-0.007	-0.070^{*}	-0.071^{*}
	(0.028)	(0.028)	(0.041)	(0.042)
Reputn. Concern		-0.003		-0.021
		(0.051)		(0.021)
Second Order Beliefs		-0.174^{***}		-0.106^{***}
		(0.045)		(0.031)
Ineq. aversion		0.048		0.040
		(0.029)		(0.034)
Language	0.132	0.116	-0.221	-0.229
	(0.159)	(0.151)	(0.184)	(0.177)
Constant	2.326^{***}	2.358^{***}	2.279^{***}	2.370^{***}
	(0.332)	(0.424)	(0.397)	(0.387)
Observations	171	171	172	172
Adjusted R^2	0.143	0.210	0.163	0.192

Note: *p<0.1; **p<0.05; ***p<0.01
HC-Robust SE in parentheses