# Should We Take Experimental Recommendations at Face-Value ? Social-Image and Self-Sorting in a Public-Good Experiment

Antoine Hémon*

*"The attempt to provide public goods without coercion is an archetypal example in which departures from pure self-interest can be beneficial to society."*

(Rabin, 1993)

## 1 Introduction

Many laboratory experiments about public goods found that subjects' contributions are initially higher than the Nash prediction, but lower than the efficient level (Ledyard, 1995). Aside from a literature studying and incorporating seemingly unselfish behaviors into economics, experimental studies search for institutional designs that foster such prosocial behaviors. For example, higher initial levels of contribution as well as a longer-standing cooperation can be observed if one introduces communication before subjects' decisions (Isaac & Walker, 1988), the possibility of sanctions – either monetary or non-monetary (Fehr & Gächter, 2000 ; Masclet, Noussair, Tucker & Villeval, 2003) or the public disclosure of individual choices to play on social-images incentives.

*Université Paris 1 Panthéon-Sorbonne & Paris School of Economics: antoine.hemon@psemail.eu

In his historical study of the incorporation of seemingly unselfish behaviors into modern economics, Fontaine (2007) traces it back to the early 1960's and concludes that the main motivations of prosocial behaviors may have been discovered at the beginning of the 1990's. If he can find early traces of invocations of image motivations being a driver for prosocial behavior (Olson, 1965 ; Arrow, 1972 ; Becker, 1974), we see as a third generation the literature based on image-motivation that only recently emerged since the 2000's as a collective investigation. Among others, Bénabou & Tirole (2006) appears as the prominent theoretical contribution to the image-motivation literature. They add a third type of motivation – an image one – being added aside with intrinsic and extrinsic (monetary) incentives. Built to rationalize previous experimental results, their model was exposed to experimental tests of its predictions, as well as theoretical extensions like Ellingsen & Johannesson (2008).

The social-image version of their model is the one that experienced the most experimental studies that unanimously lead to an increase in prosocial beahviors, especially in public-good settings. [1] Aside from previous empirical evidence of a social-image impact (Glazer & Konrad, 1996 ; Harbaugh, 1998), experimental evidence of the effect of anonymity begin with Sell & Wilson (1991) to our knowledge. Andreoni & Petrie (2004) break off with one practical standard by running "experiments without confidentiality" inside the laboratory. They propose a design using subjects' photograph to induce social-image incentives, relying on the idea that it fosters the possibility for pride and shame. They find that the mere information about individual contribution has no significant impact (contrary to Sell & Wilson, 1991) whereas the introduction of photograph does significantly increase contributions, especially when associated with individual contributions in a repeated 5-person linear VCM. This result is supported by another design proposed by Rege & Telle (2004) where the subject has to stand up with her contribution in front of the social audience in a public-good activity.

1. Of course, there is a self-signaling branch of the literature based on Bénabou & Tirole (2006) and Bodner & Prelec (2003). Yet, there is no such thing like unanimity about both its significance and its prevalence, which has some links with the "moral wiggle room" literature. See for instance Dana, Cain & Dawes (2006), Gneezy, Gneezy, Riener & Nelson (2012), van der Weele & von Siemens (2014), Grossman (2015), Grossman & van der Weele (2017).

In the same vein, Ariely, Bracha & Meier (2009) study the potential crowding-out between monetary and image incentives. They document a significant increase in effort, that induces greater donation to a charity, when the results are publicly displayed in comparison to a private-information treatment. Last, Andreoni & Bernheim (2009) questions the origins of the 50-50 norm in a dictator game by varying the probability of the dictator's choice to be implemented. As the probability decreases, they find a increasing share of dictators that share nothing and keep their whole endowment. [2]

If observability may potentially trigger various motivations, social-image incentives appear to be the one at play. Karlan & McConnell (2014) disentangle between two motivations that can drive the effect of observability, i.e. image-incentives or influencing others' behaviors. In a giving experiment to a NGO, they vary the timing of the revelation of subjects' identity, which is either before or after a last round. They do not find a significant discrepancy between both treatments whereas in presence of influencing-others motives one should have observed a higher level of donations when they are offered the possibility to influence their peers, i.e. in the treatment where the revelation come in before the last round. The evidence is even more unanimous in public-good settings. In a field setting – Lopez, Murphy, Spraggon & Stranlund (2010) – as well as in a lab experiment – Spraggon, Sobarzo & Stranlund (2015), observability by peers increases voluntary contribution in a standard linear public good games.

In these empirical studies as well as in theoretical contributions, two underlying motives can be at play : either quest-for-honor or avoiding-stigma. Samek & Sheremeta (2014) investigate these motives building on Andreoni & Petrie (AP) (2004)'s design. In a 5-person linear VCM, they vary the number of subjects whose identity will be revealed with respect to the rank of their contribution. The Baseline is a standard anonymous VCM and the full-information treatment replicates AP associating photos and contributions. They propose two additional treatments where only the two highest (High) or the two lowest (Low) contributors

2. Note that we focus on social observability by other subjects and not by the experimenter. Based on Barmettler, Fehr & Zehnder (2012), we estimate the latter as negligible for the mere identification of image-motivation, even if their results only hold for dictator, ultimatum and trust games.

have their photos displayed to others in their group. They replicate AP's previous findings and find that the Low treatment induces the same level of contributions as the full information treatment, whereas the High treatment show significantly lower contributions than both other treatments. Thus, in their settings, social image concerns seem to pass through stigma avoidance rather than search for prestige. However, the two authors find the same levels of contributions for both selective-information treatments in a framed-field experiment (Samek & Sheremeta, 2017). [3]

However, these evidence that social-image motives induce higher contributions in public-good situations should not make us to jump to its implementation outside of experimental settings. Indeed, these different underlying motives may have different impact on the participation decisions of agents who can decide whether or not to enter or to exit an environment. Rare enough to be mentioned, Samek & Sheremeta (2014, 2016) combine external recommendations about charity campaigns as well as explicit cautious requirements with respect to their results. [4] Self-sorting is a candidate to challenge the external validity of the higher efficiency of social-image environments. In particular, if avoiding the stigma is the main driver for higher contribution, some agents will not enter the mechanism if they are offered the opportunity to do so.

To date, it is not clear how far experimental economics can go to tackle this self-sorting threat (Al-Ubaydi & List, 2017). [5] In this paper, we propose to study the interaction between an explicit outside option and social-image incentives. This paper is one of the first to explicitly study the interaction between an outside option and an institutional variation meant to foster prosocial behaviors. In a double dictator game, Malmendier *et al.* (2014) study the interaction between different reciprocity contexts (either negative, neutral or positive) and a outside option that allows dictators not to participate in the sharing decision and keep the whole endowments. Regner (2017) does not find any significant interaction between an

3. The two same authors also find that the mere threat of observability can induce an increase of contributions (Samek & Sheremeta, 2016)

4. In particular, "[g]iven the opportunity of free entry and exit, individuals may simply avoid contributing to communal and charity groups that identify the lowest contributors."

5. Details about the self-sorting literature are presented in the introduction of Chapter 1.

outside option and communication.

Based on a previous experiment, we implement the same opting-out procedure where it is pay-off equivalent not to participate, and to participate to free-ride. We propose an experiment made of three treatments. The Baseline is a standard linear VCM with social-image *à la* Andreoni & Petrie (2004) – the only differences being that it is a 4-person VCM and that group photographs are displayed only on the outcome screen and not during the whole procedure. The Baseline is a benchmark for the two treatments with the outside option we implemented. The only difference between the two is that in Treatment 1, a subject that does not participate will have her photograph displayed to others whereas in Treatment 2, her photograph is not displayed. By this design variation, we are able to control for a potential fly-to-anonymity. Indeed, in our previous experiment, we find that not-to-participate in a VCM and to-participate-to-contribure-zero were the two actions that were the less socially appropriate actions in a VCM with an outside option : they were not statistically distinct and but significantly different from other positive contributions. Thus, one can anticipate that more subjects will take the outside option if it offers anonymity, i.e. in Treatment 2.

Based on this treatment structure, subjects first play a one-shot VCM followed by beliefs elicitations about others' mean contribution and participation decisions where relevant. This is supposed to enable us to make cautious comparisons with our previous studies. A third part is a 8-period repeated VCM similar to the one-shot activity to offer a primer in the dynamics associated with outside options. The experiment ends with a questionnaire.

Our results are the followings. First, we observe a high level of contribution in an environment with social-image incentives when participation is coerced. Second, one-shot exit rates range from 7% in Treatment 1 to 14% in Treatment 2. Self-sorting is significant with respect to baseline, but the difference across treatments is just above conventional significance levels, even if facially it gives weight to the fly-to-anonymity argument. Third, contrary to our Chapter 1 anonymous environment, offering an outside option does have a negative impact on mean contribution levels as well as in public-good provision. In a coerced environment, social-image institutions achieve 2/3 of the optimal provision whereas in an environment

with voluntary participation, it reaches levels around 50% of the optimum.

Fourth, the elicited belief about others' contribution appears as the prominent driver of subjects' contributions. On the one hand, the treatment effects are no longer significant in its presence, which suggests that it is the main channel to rationalize the decrease in contributions between baseline and both treatments. On the other hand, the contribution belief plays a stronger role that the belief about others' participation in the two treatments.

Last, we provide a preliminary description of the dynamics in the repeated VCM. If the coerced environment replicates the well-known contribution decay over time, it does not seem to be the case once you offer an explicit outside option to subjects.

The paper is organized as follows. Section 2 exposes the design of the experiment and its empirical implementation. Section 3 presents the results. As Section 1 introduced, Section 4 concludes.

## 2 Experiment

### 2.1 Experimental Design

The experiment is made of three treatments, being implemented between-subject. The baseline is a coerced-participation VCM with social image ; Treatment 1 is a voluntary-participation VCM with social image where subjects can choose not to participate but this particular choice is associated with her photograph as well as every other contribution choice. Treatment 2 is a voluntary-participation VCM with social image where only the photographs of participants are disclosed.

Each treatment is composed of 4 parts : a one-shot public-good game, a subjects' beliefs elicitation phase, a 8-period repeated VCM, and a post-experiment questionnaire. After general instructions, the instructions for each part was only presented before its beginning such that subjects cannot infer what will happen next, except that the experiment is calibrate to last 1 hour and is composed of 4 parts.

**Part I – A One-Shot Social-Image Voluntary Contribution Mechanism.** The three treatments share the same public-good task which is either compulsory (Baseline) or optional (Treatment 1 and Treatment 2). The public-good task is a standard one-shot linear VCM with groups of size 4. They receive an individual endowment of 10 Experimental Currency Unit (ECU) – the exchange rate being 4 ECU for 1 euro. They individually and simultaneously choose the amount they want to keep and the amount they want to dedicate to a common project, being a pure public good. Each ECU kept returns 1 ECU to its keeper ; each ECU contributed to the public good yields a social gain of 1.6 ECU that is equally shared between the 4 group members, i.e. 0.4 ECU for each. This classical incentive structure is the one of a socially dilemma where individual rationality that leads to zero contribution does not go in line with the social optimum which requires full contribution by each group member. The task description is made using this wording of keeping/dedicating to a common project and some examples are provided.

This public-good activity is associated with a common procedure to induce social-image incentives, i.e. the association of subjects' photograph with their contribution decisions. We opt to include photograph to allow subjects to identify the contributors of their groups as well as themselves being identified by others. This choice is inspired by Andreoni & Petrie (2004) who make clear that photograph is a useful tool to induce social-image incentives, avoiding potential confounding effect due to communication or face-to-face set-up. Unlike AP's procedure , we had the opportunity to use institutional photographs of our subjects. [6] This feature is likely to offer better control than taking pictures of their subjects at the beginning of their experiment. If AP and among others Samek & Sheremeta (2014) display group-members photographs during the whole decision phases, we choose to reveal them only during the result phase. This design has the advantage to focus on own-social image incentive independently of one's own group composition. [7] Subjects only see their own photograph

6. In a preliminary phase, subjects have to enter their student identification number to have her institutional photograph being displayed.

7. In particular, Ellingsen & Johannesson (2008) propose a model where the composition of the audience may induce different levels of social-image incentives, due to the weights the agent grant to the audience's

during the decision phases.

**Treatment Variations.** Contrary to the Baseline where participation is compulsory, [8] the participation to this activity is optional both in Treatment 1 and in Treatment 2. First, subjects face a participation stage where they have to choose whether or not they want to participate in this task. If a subject participates, she receives her endowment and has to enter her contribution without any information about others' participation decisions, contribution decisions being simultaneous. If a subject does not participate, she still receives a 10-ECU endowment and benefits from the group contributions to the pure public good, by its mere definition of non-excludability. Thus, we have isomorphic games across treatments and in presence of endogenous entry. Yet, a non-participant doesn't have the opportunity to choose how to allocate her endowment. She just has to confirm her choice, such that nobody can infer participation decision from click noises.

The only difference between Treatment 1 and Treatment 2 is the display of the photograph when one chooses not to participate. In Treatment 1, if a subject opts for non-participation, her photograph is displayed on every group-member output screens associated with her participation decision. In Treatment 2, for the same decision, her photograph is not displayed and replaced by a crossed-out blank. With this design variation, we aim to be able to disentangle between a pure outside option – for situations where participants can infer who is present and who is absent (e.g. small groups) – and an outside option that also provides anonymity – for situations where participants cannot infer who avoided the ask (e.g. larger groups). Indeed, Chapter 1 provides evidence that a non-participation in a VCM action was seen as socially inappropriate as a zero-contribution actions, and social-appropriateness was positively correlated with contribution levels. Thus, we can anticipate a fly to anonymity for null and low contributors.

The procedure was as follows. In a preliminary phase, subjects enter their student ID number to have their photograph displayed on their screen only. Then, subjects are presented

judgment.

8. Subjects can only click that they want to participate in the task, to have the closest treatment comparability.

the public-good activity and they are provided examples of output screens to make sure that they understand when their photograph will be displayed to their group-members. They enter their decisions seeing only their own photograph. Last, on the output screen, the photographs and associated contributions were ranked from the highest to the lowest. The photographs were not displayed only in Treatment 2 for non-participants. Note that the output screen was displayed only after Part II to perform belief elicitations.

**Part II – Subjects' beliefs elicitations.** The second part of the experiment aims at eliciting subjects' beliefs about the behaviors of their group-members.

First, in every treatment, we elicit subjects' belief about the mean contribution of the members of their own groups. A binarized scoring rule (BSR) is implemented (Hossain & Okui, 2013). Subjects are asked to guess the mean contribution of others under the following incentive-compatible scheme. The difference between the guessed mean contribution and the actual mean contribution – the guessing gap – is squared and compared to a random number that is uniformly drawn between 0 and 100. If the guessing gap is lower than the number drawn, a subject earns 2-ECU prize ; if it is larger, she gets a lower – zero ECU – prize. To make the BSR incentive-compatibility salient, the instructions explicitly told subjects that this mechanism guarantees that their best-response is to give their true belief (following Dargnies *et al.*, 2017).

Second, only in Treatment 1 and Treatment 2, this part additionally includes belief elicitation about others' participation decisions using the same BSR procedure. Subjects had to guess a number of participants between 0 and 3 and the guessing gap is compared to a number uniformly drawn between 0 and 9. This procedure yields subjects' belief about the mean number of participants in their groups, themselves excluded.

**Part III – Repeated VCM.** The third part of the experiment was made to initiate a study of the dynamics of behaviors in the presence of an outside option. It is a finitely repeated VCM identical to the one subjects went through in Part I. We opted for 8 periods due to session duration constraint as well as to match Andreoni & Petrie (2004) 8-periods

design.

Due to the fact that taking the outside option may provide subjects with anonymity, we opted for a pure stranger design, where new groups are reformed with replacement at each period. Note that this stranger design is conservative with respect to social-image since one has no incentives to engage in reputation building, contrary to both previous studies that favor repeated VCM with fixed-groups, even if A&P rematch groups after 8 rounds five times.

Our design proposes a combination between one-shot and repeated versions of the same activity. Subjects are not aware of a subsequent repeated version of the VCM when playing the one-shot VCM. We can expect a restart effect for Part III that is documented in the literature, an hypothesis we will have to investigate.

**Part IV – Post-experiment questionnaire.** The experiment ends with a questionnaire. The questionnaire first asks socio-demographic questions (age, gender, education, occupation – even if students) as well as items about previous experimental experience (they all have some since they went through the whole pedagogical track). It then gathers survey variables associated to risk aversion (Dohmen *et al.*, 2011), with trust in others with the question of the World Value Survey (Johnson & Mislin, 2012) with confusion, or more specifically one's ability to identify the Nash equilibrium of the one-shot VCM (Ferraro & Vossler, 2010) A last section consists in two hypothetical choices about time preferences and ambiguity. We also include one item about social-image that is closely correlated by the "image-concern game" proposed by Henry & Sonntag (2015).

### 2.2 Experimental Procedures

The experiment was conducted in France at LESSAC – Laboratory for Experimentation in Social Sciences and Behavioral Analysis – at the Burgundy School of Business in Dijon. We ran 18 computerized sessions between the 15th of October 2018 and the 13th of November 2017. Only 16 are workable since one session experienced a power failure and subjects went

through another treatment session. The photographs were those used for their curriculum at their institution, that can be used for pedagogical purposes such as experiments. [9]

The sessions were part of a course program about Behavioral Organizations where students have to be present at 7 experiments in the same order, each of it being followed by a debriefing session. The present experiment was the 6th session but subjects have no previous experiences with respect to public good games. 5 persons administrated the sessions associated to their students' classes. To limit experimenter effect, instructions were computerized too and subjects were asked to carefully read the instructions by themselves. [10] In two baseline sessions (4 and 7), non-fluent French-speaking students were present such that the administrator had to read and translate on-the-fly, the administrator's mother tongue being English. We collect data about 388 subjects : 140 for Baseline, 112 for Treatment 1 and 136 for Treatment 2. Sessions' size goes from 16 up to 32 subjects. Practically, subjects waited at the entrance of the laboratory for the administrator to get them on the lab. They then sit at computer stations.

The maximum duration was set at 1 hour. The average payment was 7.5 euros with no show-up fee and similar across treatments (respectively 7.8 for Baseline, 7.4 for Treatment 1 and 7.3 for Treatment 2).

## 3 Results

### 3.1 One-shot Analysis

Table 1 provides frequencies of subjects' participation decisions. We observe that 7% of subjects prefer to take the outside option rather than to participate in the social-image VCM in Treatment 1. The exit rate doubles when the outside option offers anonymity (14%). These differences in participation with respect to Baseline are significant (Fisher's Exact Test for Count Data, p-value$<$ .01 for both Baseline vs T1 and Baseline vs T2). The discrepancy in

9. The Burgundy School of Business got a Pedagogical Innovation prize in 2009 for their program that allows for academic use of students' experimental data.

10. The experimental software was z-TREE (Fischbacher, 2007).

TABLE 1 – Participation Choices across treatments

| | Don't participate | Participate |
|---|---|---|
| Baseline | 0.00 | 1.00 |
| Treatment 1 | 0.07 | 0.93 |
| Treatment 2 | 0.14 | 0.86 |

**Notes :** This table displays participation-choice frequencies across treatments.

participation across the treatments is not significant at conventional level (Fisher's Exact Test for Count Data, p-value= 0.103).

These exit rates are similar to the 16% exit rate that we find in Chapter 1. It should be understood as a floor level under the assumption that subjects who volunteer to participate in experiments come to play (Carpenter *et al.*, 2010).

We now turn to contribution patterns. Figure 1 displays the contribution behaviors of subjects who participate in the VCM. On the left panel, Figure 1.a shows the mean contributions of participants across treatments. On the right panel, Figure 1.b displays the cumulative distribution of contributions in terms of subjects initial endowments of participants.

In Baseline, subjects contribute almost 2/3 of their endowment (6.7 ECU out of their initial 10 ECU). It is higher than usual contribution levels in one-shot VCM which is coherent with the social-image hypothesis of higher contributions and the scarce experimental we have at hand. Indeed, if we compare to Andreoni & Petrie (2004)'s first-round results, the average percent of endowment contributed to the public lies between 50% and 60% whereas in their baseline (standard anonymous VCM) it lies between 30% and 40%. For Samek & Sheremeta (2014, 2016), they find a contribution rate between 60% and 70% with social-image and just under 50% for their anonymous baseline. Thus, the contribution rate we observe in our baseline are in line with the corresponding literature. Moreover, we can cautiously compare the contribution level to the one we observed in a similar anonymous VCM we run in Chapter 1. The two procedures are almost similar, [11] but the place, the pool and the date are different.

11. In the Chapter 1 experiment, subjects are provided with extra-understanding questions taken from

We obtained a mean contribution of 3.99 ECU, which is significantly lower than the 6.7 ECU mean in our current social-image Baseline (Two Sample t-test : t = −5.55, df = 234, p-value < .001). This higher contribution level on average can be explained by a low level of free-riders (9%) and a high fraction of full contributors (45%). More, we see that subjects who contributed less than half their endowment only constitute 1/4 of the population (26.4%). 15% choose a fifty-fifty allocation.

When the design offers an explicit outside option (either Treatment 1 or Treatment 2), we observe lower contributions means (respectively 5.9 ECU and 5.6 ECU). Both differences with respect to Baseline are significant (Two Sample t-test : t = 1.8857, df = 242, p-value = 0.06 for Baseline vs Treatment 1 ; Two Sample t-test : t = 2.4945, df = 255, p-value = 0.01 for Baseline vs Treatment 2). We have a significantly negative treatment effect on the mean contribution for the participants. The lower shares of free-riders (resp. 2% and 5%) is counterbalanced by the decrease in full contributors of 20 percentage points (resp. 27% and 27%). In fact, we do observe significant distribution discrepancies between Baseline and Treatments (Two-sample Kolmogorov-Smirnov tesy : D = 0.1765, p-value = 0.038 for Baseline vs Treatment 1 ; Two-sample Kolmogorov-Smirnov test : D = 0.18077, p-value = 0.04 for Baseline vs Treatment 2).

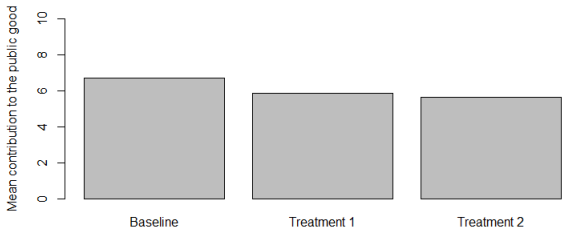
Surprisingly, we do not observe any significant differences across the two treatments with an outside option. Mean contributions are not statistically different (Two Sample t-test : t = 0.58371, df = 219, p-value = 0.56), as well as their contributions distributions (Two-sample Kolmogorov-Smirnov test : D = 0.076923, p-value = 0.90). The impact of the anonymity conferred by the outside option in Treatment 2 seems only to pass by an lower participation rate, even if not significant at conventional levels.

Note that all the comparisons between the baseline and the treatments are conservative since we do not take into account that non-participants *de facto* contribute zero to the public good, which leads to even lower mean contributions and lower global provision of the public

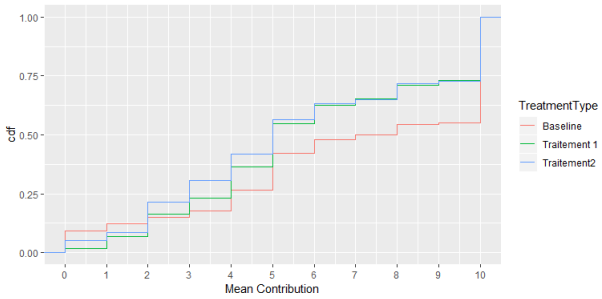
Ferraro & Vossler (2010) to reduce confusion that is supposed to lead to a fifty-fifty allocation. We then observe a similar share of subjects that allows their endowment on a 50-50 basis (11% in the anonymous VCM Baseline vs 15% in the current social-image VCM Baseline.

FIGURE 1 – Mean Contributions and Distributions of Contributions Across Treatments

1.a – Mean Contributions




1.b – Participants' contribution CDF




good that participants' mean contributions can suggest. Indeed, the ex-post contribution means are lower in Treatment 1 and Treatment 2 (resp. 5.5 ECU and 4.8 ECU) and the differences in means with respect to baseline reach higher level of significance (Two Sample t-test : $t = 2.8341$, $df = 250$, p-value $< .001$ for Baseline vs Treatment 1 ; Two Sample t-test : $t = 4.3125$, $df = 274$, p-value $< .001$ for Baseline vs Treatment 2). Thus, the provision of public good is on average 26.8 ECU in Baseline or a gap of 1/3 from the optimal provision of 40 ECU, whereas the provision levels are around 50% of the optimal provision in Treatments (resp. 55% and 48%).

We now turn to the determinants of heterogeneity in individual behaviors. Table 6 displays 3 OLS regressions where participants' contribution is the dependent variable. Column 1 only regresses participants' contribution on Treatment dummy variable. We obviously find the negative correlation between both Treatments and participants' contributions with respect to Baseline. Being offered an outside option with respect to be coerced to participate is correlated with a 1-ECU lower contribution by subjects that opted in – respectively $-0.815$ for T1 (p-value=.059) and $-1.068$ for T2 (p-value=.01).

Column 2 adds the elicited belief about the mean contribution of the others in one's group, as well as the associated interactions terms. We observe that the only significant variable is this belief variable. An expectation of a 1-ECU higher mean contribution of others is associated with an increase of .976 ECU of her own contribution. This can be explained by

TABLE 2 – OLS Regressions – Participants' Contributions to the Public Good

| | *Dependent variable :* | | |
|---|---|---|---|
| | Participants' Contribution | | |
| | (1) | (2) | (3) |
| T1 | $-0.815^{*}$ | 0.366 | 0.049 |
| | (0.431) | (0.904) | (0.914) |
| T2 | $-1.068^{**}$ | $-0.218$ | $-0.353$ |
| | (0.417) | (0.906) | (0.913) |
| Contribution Belief | | $0.976^{***}$ | $0.943^{***}$ |
| | | (0.088) | (0.090) |
| Truster | | | $0.597^{*}$ |
| | | | (0.335) |
| Risk Aversion | | | $0.707^{**}$ |
| | | | (0.317) |
| Contrib. Belief :T1 | | 0.032 | 0.073 |
| | | (0.152) | (0.153) |
| Contrib. Belief :T2 | | 0.056 | 0.065 |
| | | (0.149) | (0.149) |
| Constant | $6.700^{***}$ | 0.434 | $-1.538$ |
| | (0.281) | (0.606) | (3.297) |
| n.s. Controls | | | *Added* |
| Observations | 361 | 361 | 361 |
| $R^2$ | 0.020 | 0.436 | 0.459 |

**Notes :** OLS regressions. Standard errors in parentheses; $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01; n.s. Controls refer to insignificant control variables. They are age, gender, social image, present preferences and ambiguity survey proxies and a no-confusion dummy.

(anticipated) reciprocity. This results leads us to think that the main channel to rationalize the decrease in explicit contributions between baseline and both treatments transit through the contribution belief channel. Indeed, we observe a significant difference in means about the contribution beliefs between the baseline and both treatments (Welch Two Sample t-test : t = 4.9019, df = 249.44, p-value $< .001$ ; Welch Two Sample t-test : t = 4.9241, df = 267.43, p-value $< .001$ for Baseline vs T2), but not between the two treatments (Welch Two Sample t-test : t = -0.22538, df = 235.96, p-value = 0.8219). [12] This results holds for total expected contributions by others. Remember that during the elicitation task about others' contribution, the question was about others' mean contribution as a whole and not only participants. [13]

Column 3 just adds control variables. It does not change the magnitude and the significance of the relation between the contribution belief and participants' contribution. Note that only trust and risk aversion yields significant (and positive) correlation with participants' contribution.

Last, we focus on the different potential effect of the elicited beliefs about others' behaviors on the subsample of subjects of Treatment 1 and Treatment 2 only, since we do not have such data for Baseline. Table 3 presents the results from the 4 following regressions. The first column only focuses on the correlation between one's contribution to the public good and her Contribution belief. As we saw it previously for the 3 treatments together, we observe a significant and positive correlation between the two of a magnitude of 1 (p.value$< .001$). The second column regresses one's contribution on her belief about the expected number of participants. The correlation is also positive and significant (.86 for a p-value= .014) ; but this participation belief predicts very little of the variance of contributions ($R^2 = 2.7\%$ vs $R^2 = 39.6\%$ for the contribution belief). Indeed, conditional to contribution beliefs (column 3), this correlation between participation beliefs and contribution behaviors is reduced and

12. Note that we do not observe a significant difference in the number of expected participants in the two treatments (Welch Two Sample t-test : t = -0.40452, df = 238.21, p-value = 0.68).

13. If we use other proxies to get the total expected contributions using the belief about others' participation and using 3 for Baseline, we do not find differences in results.

TABLE 3 – OLS Regressions – Impacts of beliefs about contribution and participation

| | *Dependent variable :* | | | |
|---|---|---|---|---|
| | Participants' Contributions | | | |
| | (1) | (2) | (3) | (4) |
| Contribution Belief | 1.015*** | | 1.012*** | 0.980*** |
| | (0.085) | | (0.088) | (0.088) |
| Participation Belief | | 0.860** | 0.026 | 0.012 |
| | | (0.349) | (0.285) | (0.286) |
| Risk Aversion | | | | 1.497*** |
| | | | | (0.424) |
| Constant | 0.525 | 3.549*** | 0.469 | −1.577 |
| | (0.467) | (0.918) | (0.772) | (4.465) |
| n.s. Controls | | | | *Added* |
| Observations | 221 | 221 | 221 | 221 |
| $R^2$ | 0.396 | 0.027 | 0.396 | 0.453 |

**Notes :** OLS regressions. Standard errors in parentheses; *p<0.1; **p<0.05; ***p<0.01; n.s. Controls refer to insignificant control variables. They are age, gender, social image, trust, present preferences and ambiguity survey proxies and a no-confusion dummy.

no longer significant at conventional levels, Contribution beliefs still being positively and significantly correlated with contributions at the same level (1.012). Column 4 only includes control variables into the specification where only a risk aversion proxy enters significantly, but the Contribution belief keeps the same characteristics. These results support the argument that the belief about mean contribution is more salient than the belief about the mean number of participants induced by an outside option in standard linear VCM for a non-excludable public good. This statement echoes the Chapter 1's results. Moreover, they substantiate the idea that the potential effect of an outside option on contribution patterns is mainly driven by beliefs about others' contribution. [14]

### 3.2 Outlook of the repeated VCM

In this section, we only provide a brief description of the average behaviors in the repeated VCM part of 8 rounds. Note that we let the one-shot round as the zero round as a reminder and taking into account that subjects already played the VCM once when they began Part III.

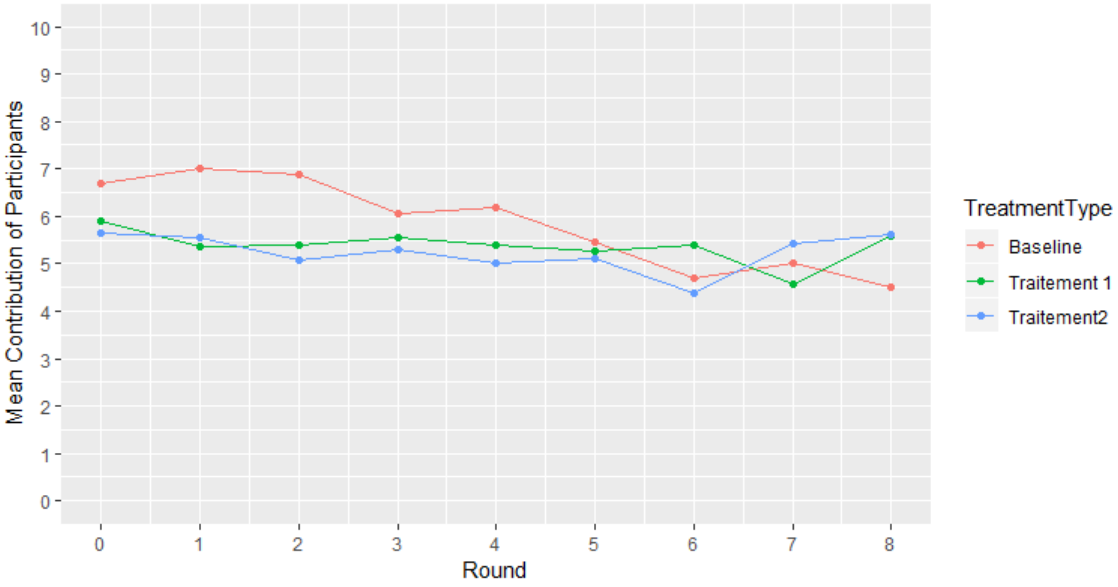
First, Figure 2 represents the means contributions across round for each treatment. Figure 2.a only refers to participants' contributions whereas Figure 2.b counts non-participants as zero-contributors since they de facto contribute zero to the public good. For Baseline, we observe a decreasing trend where the mean contribution at the first round is 70% of subjects' endowments to reach a level lower than 50% at the last period (45%). For Treatment 1 and Treatment 2, we see that participants' mean contributions vary within a range of 50-60% of initial endowments (at the exception of round 6 for Treatment 2 and round 7 for Treatment 1). For both, there is no clear pattern of decay or increase of participants' mean contribution. It is as if those who opted for participating maintain a similar degree of contribution.

When we pool participants' contributions and non-participants as zero contributors, we

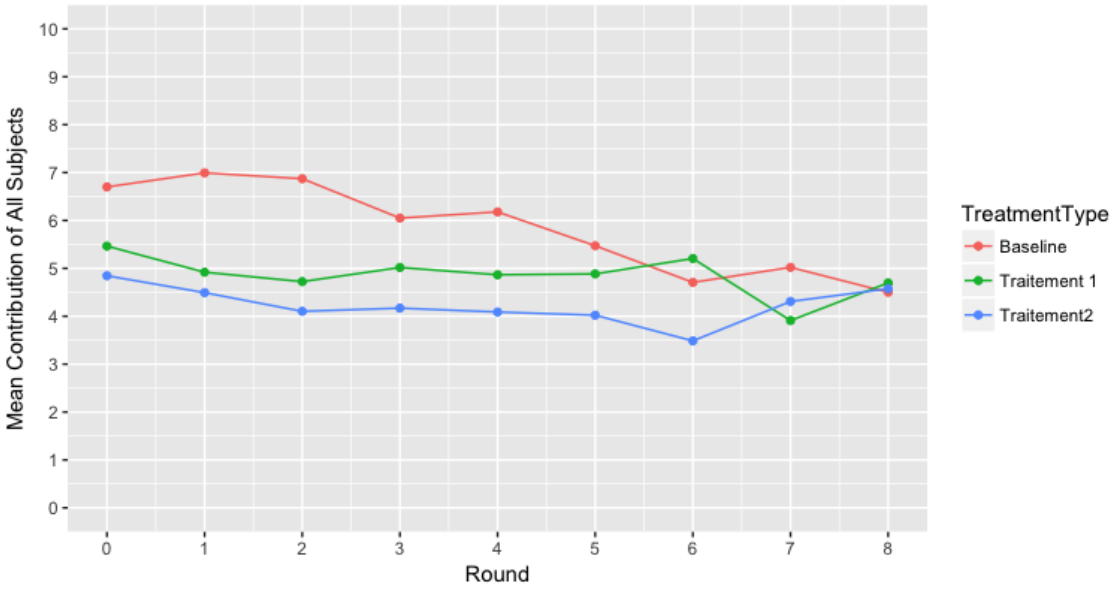
14. Note that when we take as dependent variable not participants' contributions but ex-post contributions, i.e. counting non-participants as zero-contributors, the participation-effect relation remains significant conditional on contribution beliefs (0.800, p-value=0.005), but not anymore with controls (0.012, p-value=0.97).

FIGURE 2 – Mean Contributions across Rounds per Treatments

2.a – Mean contributions of participants across rounds




2.b – Mean contributions of all subjects across rounds




trivially observe lower mean contributions for both treatments with respect to their participants' subsample counterparts. Yet, we still cannot observe a decaying trend in any of the opt-out treatments.

TABLE 4 – Summary statistics of contribution and participation across rounds

| | Round | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 0(ne-shot) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Global Mean | 6.12 | 6.07 | 5.89 | 5.68 | 5.59 | 5.30 | 4.82 | 5.02 | 5.15 |
| Baseline Mean | 6.70 | 6.99 | 6.87 | 6.05 | 6.18 | 5.47 | 4.71 | 5.02 | 4.50 |
| T1 Mean | 5.88 | 5.35 | 5.40 | 5.56 | 5.40 | 5.26 | 5.40 | 4.56 | 5.60 |
| T2 Mean | 5.63 | 5.55 | 5.07 | 5.30 | 5.01 | 5.11 | 4.39 | 5.43 | 5.60 |
| BvsT1 - Mean Diff. | 0.82 | 1.64 | 1.47 | 0.49 | 0.78 | 0.21 | -0.69 | 0.46 | -1.10 |
| (p-value) | (0.06) | (0.00) | (0.00) | (0.37) | (0.14) | (0.69) | (0.19) | (0.41) | (0.06) |
| BvsT2 - Mean Diff. | 1.07 | 1.44 | 1.80 | 0.75 | 1.17 | 0.36 | 0.32 | -0.40 | -1.10 |
| (p-value) | (0.01) | (0.00) | (0.00) | (0.17) | (0.03) | (0.52) | (0.56) | (0.47) | (0.05) |
| T1vsT2 - Mean Diff. | 0.25 | -0.21 | 0.33 | 0.27 | 0.39 | 0.15 | 1.01 | -0.86 | -0.01 |
| (p-value) | (0.56) | (0.64) | (0.50) | (0.61) | (0.44) | (0.77) | (0.05) | (0.10) | (0.99) |
| Exit Rate - T1 | 0.07 | 0.08 | 0.12 | 0.10 | 0.10 | 0.07 | 0.04 | 0.14 | 0.16 |
| Exit Rate - T2 | 0.14 | 0.19 | 0.19 | 0.21 | 0.18 | 0.21 | 0.21 | 0.21 | 0.18 |

Table 4 provides summary statistics of participants' mean contributions and exit rates across round. First, pooling participants from all treatments, we observe no difference between the one-shot and the first round (resp. 61.2% and 60.7%), but a progressive decay in mean contribution towards 51.5% for the last round. Second, we have the details round by round of the mean contribution patterns that we have just exposed. Third, if we focus on mean differences across treatment, there is no significant differences in mean between the two opting-out treatments (except for round 6 where T1 shows higher mean contribution of 1 ECU at a 5%-significance level and round 7 where the difference goes in the other direction at marginal significance level). With respect to Baseline, we start from a significantly higher mean contribution of participants for both treatments and then observe a convergence where the means are not significantly distinguishable (starting at round 3 for the two trea tments, even if round 4 goes back to a significantly positive mean difference for Treatment 2). We

even see significantly lower means for the treatments in comparison to Baseline for the last round (-1.1 ; p-value=.06 for Treatment 1, p-value=.05 for Treatment 2). This last result may be induce by a last round effect that would be different between Baseline on the one hand and the two opting-out treatments on the other hand.

Taking into account that previous contribution means only take into account participants, we now describe exit rates in both Treatments – since no exit was allowed in Baseline. In Treatment 2, exit rates remains quite constant during the repeated VCM. In the first round, 19% of subjects prefers to opt-out rather than to participate in the VCM and it then fluctuates between a floor of 18% and a ceiling of 21%. In Treatment 1, the exit rate starts at 8% and seems to increase until it doubles in the last round. Yet, the lower rates in rounds 5 and 6 (resp. 7% and 4%) prevent us to have a clear pattern to deal with.

## 4 Conclusion

This paper proposes to assess the robustness of social-image designs to foster prosocial behaviors in a public-good context with an explicit outside option. Since social-image incentives may be driven by incentives based on avoiding stigma associated to some actions, an outside option may be used to exit a mechanism designed to induce higher contributions, that would be made by reluctant contributors.

In a coerced environment, we replicate the pattern of high contributions both in one-shot and repeated VCM. Yet, when offering an outside option to subjects, we observe a lower level contribution on average with an exit rate ranging from 7% when the outside option does yield anonymity to 14% when it does, giving marginally significant evidence of the fly-to-anonymity hypothesis in a one-shot environment. Offering an outside option does significantly decrease both the individual contributions of participants and the global provision of the public good. Still, provision levels remains at levels around 50% of the optimal provision of the public good (w.r.t 2/3 when participation is coerced), which is higher than what we observed in the anonymous environments of our Chapter 1. Note that we do not observe any significant differences in contribution between Treatment 1 and Treatment 2

when one opts in.

The main driver of these treatment effects appears to be subjects' beliefs about others' mean contribution. In particular, in both treatments, when controlling for the belief about others' participation, the contribution belief remains significant and of the same magnitude whereas the participation belief is no longer significant. Contribution saliency may be at stake, even if this just indirect evidence of it.

In repeated interactions, we observe a stable exit rate when the outside option yields anonymity and an increasing but erratic exit rate when participation decisions are disclosed to the other members of the group. If we observe the standard decay in mean contribution across rounds in the coerced environment, it is far less pronounced when an outside option is at hand for subjects. It is unexpected since non-participation induces *de facto* zero contributions to the public good. It may be due to a floor level of contributions for the subjects who are willing to participate in a social-image environment.

All in all, if self-sorting is correctly captured – as a floor level – by our procedure, taking into account this threat to the validity of social-image recommendations that modify the incentives at play in the public-good mechanism does not make the social-image effect disappear, but it reduces its impact. Thus, the generalization of such self-sorting tests for studies that aimed not only at fundamental research but also at external recommendations is a safeguard. A safeguard not to implement designs that may backfire when the opportunity of exit exists. A safeguard not to over-sell lab-results to decision makers and thus to maintain experimental economics credibility. Thus, a safeguard to keep the ears of princes to whisper in.

## Appendix

TABLE 5 – OLS Regressions – Average Treatment Effects pooling an Anonymous VCM and a Social-image VCM with and without an outside option

| | *Dependent variable :* | | | |
|---|---|---|---|---|
| | Participants' Contributions | | Ex-post Contributions | |
| | (1) | (2) | (3) | (4) |
| Social Image (SI) | 2.710*** | 2.710*** | 2.710*** | 2.710*** |
| | (0.466) | (0.471) | (0.477) | (0.486) |
| Opt-out_T1 | 0.629 | | −0.110 | |
| | (0.526) | | (0.514) | |
| SI∗Opt-out_T1 | −1.445** | | −1.126 | |
| | (0.696) | | (0.687) | |
| Opt-out_T2 | | 0.629 | | −0.110 |
| | | (0.531) | | (0.524) |
| SI∗Opt-out_T2 | | −1.697** | | −1.745** |
| | | (0.694) | | (0.685) |
| Constant | 3.990*** | 3.990*** | 3.990*** | 3.990*** |
| | (0.359) | (0.363) | (0.367) | (0.374) |
| Observations | 424 | 437 | 448 | 472 |
| $R^2$ | 0.088 | 0.081 | 0.100 | 0.091 |

**Notes :** OLS regressions. Standard errors in parentheses ; *p<0.1 ; **p<0.05 ; ***p<0.01 ; We pooled data from the 2 experiments of Chapters 1 and 2. The independent variables are treatment dummies. The dependent variables are either participants' contributions or ex-post contributions.

**Références**


[1] Omar Al-Ubaydli, John A. List, and Dana L. Suskind. What Can We Learn from Experiments ? Understanding the Threats to the Scalability of Experimental Results. *American Economic Review*, 107(5) :282–286, May 2017.

[2] James Andreoni and B. Douglas Bernheim. Social Image and the 50–50 Norm : A Theoretical and Experimental Analysis of Audience Effects. *Econometrica*, 77(5) :1607–1636, 2009.

[3] James Andreoni and Ragan Petrie. Public Goods Experiments Without Confidentiality : A Glimpse Into Fund-Raising. *Journal of Public Economics*, 88(7-8) :1605–1623, 2004.

[4] Dan Ariely, Anat Bracha, and Stephan Meier. Doing Good or Doing Well ? Image Motivation and Monetary Incentives in Behaving Prosocially. *American Economic Review*, 99(1) :544–555, March 2009.

[5] Kenneth J. Arrow. Gifts and Exchanges. *Philosophy & Public Affairs*, 1(4) :343–362, 1972.

[6] Franziska Barmettler, Ernst Fehr, and Christian Zehnder. Big Experimenter Is Watching You ! Anonymity and Prosocial Behavior in the Laboratory. *Games and Economic Behavior*, 75(1) :17–34, May 2012.

[7] Gary S. Becker. A Theory of Social Interactions. *Journal of Political Economy*, 82(6) :1063–1093, 1974.

[8] Ronit Bodner and Drazen Prelec. Self-Signaling and Diagnostic Utility in Everyday Decision Making. *The psychology of economic decisions*, 1 :105–26, 2003.

[9] Boulu-Reshef, Béatrice and Jeremy Hutchison-Krupat. Leadership Communication, Transparency, and Skin-in-the-Game. *Mimeo Paris 1*, 2018.

[10] Roland Bénabou and Jean Tirole. Incentives and Prosocial Behavior. *American Economic Review*, 96(5) :1652–1678, December 2006.

[11] Colin F. Camerer, Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan, Emma

Heikensten, Felix Holzmeister, Taisuke Imai, Siri Isaksson, Gideon Nave, Thomas Pfeiffer, Michael Razen, and Hang Wu. Evaluating Replicability of Laboratory Experiments in Economics. *Science*, 351(6280) :1433–1436, March 2016.

[12] Jason Dana, Daylian M. Cain, and Robyn M. Dawes. What You Don't Know Won't Hurt Me : Costly (But Quiet) Exit in Dictator Games. *Organizational Behavior and Human Decision Processes*, 100(2) :193–201, July 2006.

[13] Marie-Pierre Dargnies, Rustamdjan Hakimov, and Dorothea Kübler. Self-Confidence and Unraveling in Matching Markets. Discussion Papers, Research Unit : Market Behavior SP II 2016-210, Social Science Research Center Berlin (WZB), 2016.

[14] Thomas Dohmen, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. Individual Risk Attitudes : Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association*, 9(3) :522–550, June 2011.

[15] Tore Ellingsen and Magnus Johannesson. Pride and Prejudice : The Human Side of Incentive Theory. *American Economic Review*, 98(3) :990–1008, June 2008.

[16] Ernst Fehr and Simon Gächter. Cooperation and Punishment in Public Goods Experiments. *The American Economic Review*, 90(4) :980–994, 2000.

[17] Paul Ferraro and Christian Vossler. The Source and Significance of Confusion in Public Goods Experiments. *The B.E. Journal of Economic Analysis & Policy*, 10(1) :1–42, 2010.

[18] Urs Fischbacher. z-Tree : Zurich Toolbox for Ready-Made Economic Experiments. *Exp Econ*, 10(2) :171–178, June 2007.

[19] Philippe Fontaine. From Philanthropy to Altruism : Incorporating Unselfish Behavior into Economics, 1961-1975. *History of Political Economy*, 39(1) :1–46, March 2007.

[20] Amihai Glazer and Kai A. Konrad. A Signaling Explanation for Charity. *The American Economic Review*, 86(4) :1019–1028, 1996.

[21] Ayelet Gneezy, Uri Gneezy, Gerhard Riener, and Leif D. Nelson. Pay-What-You-Want, Identity, and Self-Signaling in Markets. *PNAS*, 109(19) :7236–7240, May 2012.

[22] Zachary Grossman. Self-Signaling and Social-Signaling in Giving. *Journal of Economic Behavior & Organization*, 117 :26–39, September 2015.

[23] Zachary Grossman and Joël J. van der Weele. Self-Image and Willful Ignorance in Social Decisions. *Journal of the European Economic Association*, 15(1) :173–217, February 2017.

[24] William T. Harbaugh. The Prestige Motive for Making Charitable Transfers. *The American Economic Review*, 88(2) :277–282, 1998.

[25] Tanjim Hossain and Ryo Okui. The Binarized Scoring Rule. *Rev Econ Stud*, 80(3) :984–1001, July 2013.

[26] R. Mark Isaac and James M. Walker. Communication and Free-Riding Behavior : The Voluntary Contribution Mechanism. *Economic Inquiry*, 26(4) :585–608, October 1988.

[27] Noel D. Johnson and Alexandra Mislin. How Much Should We Trust the World Values Survey Trust Question ? *Economics Letters*, 116(2) :210–212, August 2012.

[28] Dean Karlan and Margaret A. McConnell. Hey Look at Me : The Effect of Giving Circles on Giving. *Journal of Economic Behavior & Organization*, 106 :402–412, October 2014.

[29] John Ledyard. Public Goods : A Survey of Experimental Research. Working Paper 861, California Institute of Technology, Division of the Humanities and Social Sciences, 1995.

[30] Maria Claudia Lopez, James J. Murphy, John M. Spraggon, and John K. Stranlund. Comparing the Effectiveness of Regulation and Pro-Social Emotions to Enhance Cooperation : Experimental Evidence from Fishing Communities in Colombia. *Economic Inquiry*, 50(1) :131–142, 2012.

[31] Ulrike Malmendier, Vera te Velde, and Roberto Weber. Rethinking Reciprocity. *Annual Review of Economics*, 6(1) :849–874, 2014.

[32] David Masclet, Charles Noussair, Steven Tucker, and Marie-Claire Villeval. Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism. *American Economic Review*, 93(1) :366–380, March 2003.

[33] Mancur Olson. *The Logic of Collective Action : Public Goods and the Theory of Groups*. Harvard University Press, Cambridge, Mass., 1965.

[34] Matthew Rabin. Incorporating Fairness into Game Theory and Economics. *American Economic Review*, 83(5) :1281–1302, 1993.

[35] Mari Rege and Kjetil Telle. The Impact of Social Approval and Framing on Cooperation in Public Good Situations. *Journal of Public Economics*, 88(7) :1625–1644, July 2004.

[36] Tobias Regner. Reciprocity under Moral Wiggle Room : Is it a Preference or a Constraint ? *Exp Econ*, pages 1–14, November 2017.

[37] Anya Samek and Roman M. Sheremeta. Recognizing Contributors : An Experiment on Public Goods. *Exp Econ*, 17(4) :673–690, December 2014.

[38] Anya Samek and Roman M. Sheremeta. When Identifying Contributors is Costly : An Experiment on Public Goods. *Southern Economic Journal*, 82(3) :801–808, 2016.

[39] Anya Samek and Roman M. Sheremeta. Selective Recognition : How To Recognize Donors To Increase Charitable Giving. *Economic Inquiry*, 55(3) :1489–1496, July 2017.

[40] Jane Sell and Rick K. Wilson. Levels of Information and Contributions to Public Goods. *Soc Forces*, 70(1) :107–124, September 1991.

[41] John M. Spraggon, Lucía Andrea Vergara Sobarzo, and John K. Stranlund. A Note on Stochastic Public Revelation and Voluntary Contributions to Public Goods. *Economics Letters*, 126 :144–146, January 2015.

[42] Joel J. van der Weele and Ferdinand von Siemens. Bracelets of Pride and Guilt ? An Experimental Test of Self-Signaling in Charitable Giving. SSRN Scholarly Paper ID 2411441, Social Science Research Network, Rochester, NY, February 2014.

TABLE 6 – OLS Regressions – Participants' Contributions to the Public Good

| | Participants' Contribution | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| T1 | $-0.815^{*}$ | 0.366 | 0.049 |
| | (0.431) | (0.904) | (0.914) |
| T2 | $-1.068^{**}$ | $-0.218$ | $-0.353$ |
| | (0.417) | (0.906) | (0.913) |
| Contribution Belief | | $0.976^{***}$ | $0.943^{***}$ |
| | | (0.088) | (0.090) |
| Truster | | | $0.597^{*}$ |
| | | | (0.335) |
| Risk Aversion | | | $0.707^{**}$ |
| | | | (0.317) |
| Contrib. Belief :T1 | | 0.032 | 0.073 |
| | | (0.152) | (0.153) |
| Contrib. Belief :T2 | | 0.056 | 0.065 |
| | | (0.149) | (0.149) |
| Constant | $6.700^{***}$ | 0.434 | $-1.538$ |
| | (0.281) | (0.606) | (3.297) |
| n.s. Controls | | | *Added* |
| Observations | 361 | 361 | 361 |
| $R^2$ | 0.020 | 0.436 | 0.459 |

**Notes :** OLS regressions. Standard errors in parentheses; $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01; n.s. Controls refer to insignificant control variables. They are age, gender, social image, present preferences and ambiguity survey proxies and a no-confusion dummy.