Social image manipulation and the participation constraint

Roberto Galbiati* Emeric Henry† Nicolas Jacquemet‡ Itzhak Rasooly§

1 Introduction

A large literature stresses the importance of social image concerns. Theoretical work (e.g Bénabou and Tirole, 2006) emphasises that individuals may change their behaviour in order to appear more 'pro-social', and that this signalling motive persists even when observers understand that individuals are engaging in strategic image manipulation. A large number of experiments deliver a similar message: making behaviour visible (and thus activating image signalling motives) often enhances pro-social behaviour. Taken together, these two strands of literature suggest that making behaviour visible, say through public scoreboards, can have a powerful impact on behaviour.

Although visibility can make behaviour 'more prosocial' in some cases, this is less clear in settings where individuals can choose whether to participate in the scheme in the first instance. For example, imagine that a volunteering organisation chooses to introduce a performance scoreboard. While this could increase average performance per volunteer, it could deter volunteers from joining the organisation in the first place (e.g., if volunteers are concerned that they will perform poorly and thus be publicly shamed). In theory, this adverse 'participation effect' could outweigh the beneficial impact of imposing visibility on effort incentives.

In this project, we explore the 'participation effect' using both theoretical and experimental methods. In this pre-registration, we focus on our planned experiment.

2 Experimental design

Platform and sample. In total, we plan to recruit 1,250 subjects using the Prolific platform. Participants will be restricted to US residents with a Prolific approval rating of at least 98% who are using a laptop or desktop computer with a working camera.

Treatments. The basic idea is to consider the impact of imposing visibility on performance both with and without an 'exit option' for participants (which imposes a 'participation constraint'). Accordingly, we consider four treatments:

1. No disclosure/no exit
2. Disclosure/no exit

*CNRS, Sciences Po & CEPR.
†Sciences Po & CEPR.
‡Paris School of Economics & Université Paris 1 Panthéon-Sorbonne.
§Sciences Po & Paris School of Economics.

3. No disclosure/exit
4. Disclosure/exit

We will also include a fifth treatment (Disclosure/visible exit) which provides participants with an exit option but makes the act of exiting visible to others. Subjects are assigned to treatments with equal probabilities (1/5). We outline each of the treatments in much more detail below.

[0] Consent form. In all treatments, participants are given some basic information about the study via the study description on Prolific. In particular, they are told that 'You may also need to take a photo of yourself using a working laptop or desktop camera.' If they click the study link on Prolific, subjects are then given further information about the study via a consent form. The consent form again emphasises that subjects may need 'to take a picture of [themselves] that could be displayed to other participants'.

[1] No disclosure/no exit. We now outline the No Disclosure/no exit treatment in some detail; this can be thought of as the baseline case.

Demographics. After completing the consent form, subjects are asked a series of demographic questions. In particular, they are asked for: their level of education, their race/ethnicity, how many studies they have completed on Prolific, their monthly household income, and whether they have donated money to charity in the last month and year. As discussed below, these measures will be used to estimate each subject's 'type'.

Click for charity. Subjects are then told that they will conduct a simple button pressing task to raise money for the Against Malaria Foundation (AMF). To raise money for the foundation, subjects need to press the 'a' and 'b' buttons on the keyboard in an alternating fashion. The faster subjects press the buttons, the more money they will raise for AMF. Subjects are then given a practice round to familiarise themselves with the task. They are also told that they will be able to verify that the donations to AMF have been made on their behalf by clicking a link that will be sent to them after the experiment has concluded.

Before completing the task, subjects are given a standard attention check: as discussed below, this will help us to screen out inattentive subjects from the sample. They then complete the main task. Throughout the task, subjects can see how much they have raised for AMF so far; and subjects are free to end the task early at any time.

In summary, the 'baseline treatment' (no disclosure/no exit) is very straightforward: subjects are first given some demographic questions and then complete the 'click for charity' task.

[2] Disclosure/no exit. The no disclosure/exit treatment is the same as the baseline except that subject choices are visible. This allows us to explore the impact of visibility on effort holding the subject pool fixed. More precisely, subjects go through the following sequence:

Demographics. Subjects are first asked the same set of demographic questions (see above).

Visibility. Subjects are told that the amount of money that they raise in the 'click for charity' task will be shared with a group of participants whose only role is to observe their

performance. They are told that members of this group will see: their picture, their score in the button-pressing task, and the amount of money they have raised for AMF. Subjects then need to take a photo of themselves using their webcam.

Image payoffs. To investigate how subjects feel about visibility, we measure participants' willingness to pay to secure or avoid disclosure using the same procedure as Butera, Metcalfe, Morrison, and Taubinsky (2022). Briefly, subjects are given some information about 25th percentile, median, and 75th percentile of performance from a previous study. They are then asked if they would prefer anonymity or disclosure conditional on raising \$0-\$0.40 for AMF, \$0.40-\$0.80 for AMF, \$0.80-\$1.20 for AMF and so on and so forth up until \$3.20 or more. Finally, they are asked how much they would be willing to pay (out of a \$10 budget that they only receive with some probability) to enforce their choice. Truthful reporting is incentivised using Butera et al. (2022)'s procedure.

Click for charity. Subjects then complete the 'click for charity task' (see above), this time with a photo of them visible above their score.

[3] No disclosure/exit. This treatment is the same as the baseline treatment except that subjects can skip the click for charity task if they wish to — this is meant to model the situation in which subjects do not need to participate in any scheme unless they want to. Thus:

Demographics. The survey begins with the standard demographic questions.

The skip option. After the click for charity task has been explained (and practiced), subjects are told that they can skip the task entirely if they wish to. If they choose to skip the task, they reach the end of the survey.

Click for charity. If subjects choose not to skip, they then complete the click for charity task.

[4] Disclosure/exit. This treatment varies Disclosure/no exit to allow for a skipping option. Thus:

Demographics. Subjects are first asked the same set of demographic questions (see above).

Visibility. Subjects are told that the amount of money that they raise will be visible (see above).

Image payoffs. We then measure participants' willingness to pay to secure or avoid disclosure (see above).

The skip option. Subjects are then given the opportunity to skip the click for charity task (see above).

Click for charity. If they have chosen not to skip the task, subjects then complete the click for charity task.

[5] Disclosure/visible exit. This treatment is exactly the same as Disclosure/exit except for one difference. Specifically, when given the skipping option, subjects are told that 'If you skip the task, then your score will be recorded as N/A when we share it (along with your photo) with others.' Thus, in the treatment, failing to participate is not a means of avoiding potentially negative inferences.

Exact survey wording. While the overview above outlines the treatments in some detail, it does not provide the exact question wording. The exact question wording can be viewed using this link.

3 Analysis plan

Exclusion criteria. We begin by dropping any participants whose responses suggest that they have failed to pay attention to the survey. Specifically, we drop any participants who fail the attention check and drop any participants who state that they have given money to charity in the last month but not in the last year (which is impossible).

Effect of disclosure on participation rates. We then study whether disclosure reduces participation, as one might expect. To do this, we will regress the decision to participate on dummy variables indicating treatment assignment (restricting attention to the treatments in which subjects have a choice about whether to participate or not).

Effect of disclosure on effort amongst participants. We then study whether disclosure increases effort conditional on participation, as one would also expect. To do this, we will regress effort (i.e. the amount of money a participant raises) on dummy variables indicating treatment assignment, restricting attention to those who participated in the click for charity task.

Effect of disclosure on effort. If disclosure reduces participation rates but increases effort amongst participants, it is unclear whether it increases effort overall. To test this, we will again regress effort on dummy variables indicating treatment assignment, this time running the regression on the full sample (and coding non-participants as exerting zero effort).

Estimating the types of those who participate. In theory, one would expect those who choose to participate to be 'higher types' (i.e. individuals who would have put in more effort even in the absence of visibility incentives). To test this, we will follow a three-step procedure. In the first step, we will use our demographic variables to estimate the link between demographics and effort, using data only from the No Disclosure/no exit treatment. In the second step, we will use the estimated statistical model to estimate the type of all subjects, including the types of subjects in other treatments. In the third step, we will examine how estimated types vary with the decision to participate.

Estimating the payoffs generated by image concerns. Finally, we will estimate how the willingness to pay for disclosure varies with expectations about performance and whether this is concave as found by Butera et al. (2022).

References

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

Luigi Butera, Robert Metcalfe, William Morrison, and Dmitry Taubinsky. Measuring the welfare effects of shame and pride. *American Economic Review*, 112(1):122–168, 2022.