



**HAL**  
open science

## Do image spillovers deter rule breaking?

Rémi Suchon, Daniel Houser

► **To cite this version:**

| Rémi Suchon, Daniel Houser. Do image spillovers deter rule breaking?. 2018. halshs-01876569

**HAL Id: halshs-01876569**

**<https://shs.hal.science/halshs-01876569v1>**

Preprint submitted on 18 Sep 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

WP 1819 – September 2018

## Do image spillovers deter rule breaking?

Rémi Suchon, Daniel Houser

### Abstract:

We test whether individuals internalize the effects that their behavior may have on the social image of their group. In our experiment, we recruit pairs of real-life friends and study whether rule breaking in the form of misreporting decreases when misreporting may have negative spillovers on the image of the friend. We find that participants hurt their friends' social image by misreporting because external observers update their beliefs: they rightfully expect that a participant whose friend misreported is likely to misreport himself. However, participants misreport as often when their behavior can hurt the friend's image as when it cannot, even though hurting their friends' image reduces their own monetary gains. Our interpretation is that they underestimate the impact of their behavior on external observers' beliefs about their friends. Our results cast doubts on the capacity of groups to sustain a good image absent the possibility of punishment, which is bad news. The good news is that external observers may use image spillovers to update their beliefs and interact with members of social groups more efficiently.

### Keywords:

Social image, Group image, Misreporting, Experiment

### JEL codes:

C92, D91

# Do image spillovers deter rule breaking?

Rémi Suchon\* - Daniel Houser†

July 25, 2018

## Abstract

We test whether individuals internalize the effects that their behavior may have on the social image of their group. In our experiment, we recruit pairs of real-life friends and study whether rule breaking in the form of misreporting decreases when misreporting may have negative spillovers on the image of the friend. We find that participants hurt their friends' social image by misreporting because external observers update their beliefs: they rightfully expect that a participant whose friend misreported is likely to misreport himself. However, participants misreport as often when their behavior can hurt the friend's image as when it cannot, even though hurting their friends' image reduces their own monetary gains. Our interpretation is that they underestimate the impact of their behavior on external observers' beliefs about their friends. Our results cast doubts on the capacity of groups to sustain a good image absent the possibility of punishment, which is bad news. The good news is that external observers may use image spillovers to update their beliefs and interact with members of social groups more efficiently.

*Keywords:* Social image, Group image, Misreporting, Experiment.

*JEL codes:* C92, D91.

---

\*Corresponding author, Univ Lyon, ENS de Lyon, GATE UMR 5824, F-69342 Lyon, France. suchon@gate.cnrs.fr

†Department of Economics and the Interdisciplinary Center for Economic Science, George Mason University, Fairfax, VA, United States. dhouser@gmu.edu

# 1 Introduction

Because information on individuals is imperfect, individuals are often judged as members of social groups rather than as individuals (Phelps, 1972; Arrow, 1973; Aigner and Cain, 1977). When it is the case, the image of the group is substituted for the less available individual information and determines the members' economic success. Some social groups have a positive image, which benefits the members. For instance, religious people tend to be perceived as more trustworthy (Tan and Vogel, 2008). On the other edge of the spectrum, some groups have a negative image. In these cases, the members may suffer from the negative stereotypes attached to the group and be excluded from profitable economic interactions. For instance, in an experiment with Israeli participants, Fershtman and Gneezy (2001) show that eastern Jews are believed to be less trustworthy than Ashkenazic Jews (which turns out to be erroneous). Similarly, Cohn et al. (2014) show that bankers are perceived to be less honest than people from the general population and about as honest as inmates. Even more problematically, a bad image at the group level can be erroneous and self-fulfilling (Glover et al., 2017).

Because the image of social group is a major determinant of the members' economic opportunities and success (e.g. Fershtman and Gneezy, 2001; Tan and Vogel, 2008; Doleac and Stein, 2013) understanding the way group members participate in building it is important. However, little is known about this precise issue. The image of social groups have been studied mainly through the lens of stereotypical thinking (e.g. Fershtman and Gneezy, 2001; Castillo and Petrie, 2010; Bordalo et al., 2016). In this literature, groups generally reflect an image that is rather exogenous and only a few papers have examined how the concern for the image of the group impacts the behavior of its members. Tirole (1996) shows theoretically that groups are confronted with a public good problem when it comes to building a collective reputation. Huck and Lünser (2010) show experimentally that groups can sustain a somewhat correct reputation: members forego selfish gains in order to preserve the image of the group. Relatedly, Eriksson et al. (2017) show that individuals are also willing to forego monetary gains to save an ingroup member from losing face, even when group identity is minimal.

On the other hand, an ever growing literature has shown that preserving a good individual image is an important driver of behavior. Experimental literature generally finds that social exposure increases norm-obedience. When exposed, people are more altruistic (e.g. Andreoni and Bernheim, 2009), reciprocal (e.g. Bao et al., 2018) and honest (e.g. Kajackaite et al., 2018; Abeler et al., 2018). This is so because individuals associate a consumption value to having a good image and anticipate that a good image will help them in their future interactions with others. An open question is whether the well-identified impact of image concern at the individual level translates to the group level. It is logically appealing to expect group members to try to preserve their group image. One could think that group members should prefer the group to reflect a positive image to outgroups *per se*. Similarly, to the extent that group members care about their ingroups, they should be concerned about the cost a poor group image would cast on them. Thus, the question we address in this paper is: Do group members internalize the effect that their behavior might have on the image of the group ?

By answering the previous question, the present paper contributes to give micro-foundations to group image and identifies a mechanism by which stereotypical discrimination may arise. The premise of this mechanism is that, within groups, individuals share some underlying characteristics because of influence and selection. As a consequence, an external observer may use the first-hand information he has on one member of a group in order to decide how to interact with another member of the group. We term this mechanism “image spillovers”: the behavior of the observed one might spillover to the image of the second one and (in part) determine his economic opportunities. Member of groups might be concerned by the image of the group *per se* or because it determines the members’ economic opportunities. Consequently, anticipating that they might impact the image of the group could push them to behave better.

We study experimentally this mechanism in a setting in which the behavior of an individual might harm his ingroups’ image, and, as a result, their chance of success in further interactions. More specifically, we test experimentally whether image spillovers in pairs of participants deter misreporting. The structure of the game, which is common knowledge, is the following. First, both members of the

pair have to report the outcome of a random variable independently and without communication. They can report truthfully or misreport. Generally, misreporting pays more. Second, an external observer (the Audience), observes the report of one predetermined member of the pair, and then has to decide whether he is willing to interact with the other member of the pair. The Audience wants to play only if he expects that the probability of a truthful report by the second member is high enough. Both members of the pair prefer the Audience to play, because when the Audience plays, their payoff are higher.

We implement two experimental conditions, between which we vary whether image spillovers can arise across the members of the pair. In the baseline, in each pair, members of the pair do not know each other and were invited independently. In the treatment, the members of each pair are friends, and were invited jointly. The intuition is that the behavior of a friend is informative about the behavior of the other friend. The fact that friends tend to be similar echoes the popular saying “Bird of a feather flock together” and has received a large support in economics (e.g. Leider et al., 2009; Goeree et al., 2010) as well as other social sciences (e.g. Kandel, 1978; McPherson et al., 2001; Lee et al., 2009) and even biology (e.g. Christakis and Fowler, 2014). This reflects both homophily (the tendency to become friends with similar individuals) and influence (the tendency of friends to exert influence on each others). In the treatment, we expect the Audience to pull information from the behavior of the first friend and to condition his decision to play with the second friend on this information. We do not expect such a pattern in the baseline. We expect friends whose report are observed to anticipate this and to misreport less often in the treatment than in the baseline. The rational is to try and convince the Audience to play. To sum up, we expect image spillovers to arise in the treatment but not in the baseline, and the perspective of image spillovers to deter misreporting.

We find that image spillovers do not deter misreporting: participants whose reports are observed are equally likely to misreport in both conditions. However, in the treatment, the Audience uses the observation of the first member of the pair to decide whether to play or not. This behavior corresponds to the belief that friends are more likely to take the same decision than two participants taken randomly.

These beliefs are in turn corroborated by the data. By analyzing the participants' beliefs, we discuss potential explanation for this. We argue that participants whose reports are observed typically underestimate the impact their behavior has on the decision of the Audience.

The remaining of the paper is organized as follows. In section 2, we present the experimental design. In section 3, we introduce the conceptual framework and the hypotheses to be tested. In section 4, we present our results. Last, section 5 concludes.

## 2 Experimental design

### 2.1 The experimental game

The experimental game aims at capturing the impact a group member might have on his ingroups' economic opportunities through the impact of his own behavior on the image of his group.

The game is played for four periods. Before the first period, the program assigns their roles to participants. They are either Audience or member of a pair. In each pair, one participant is given the role of  $f_1$  and the other is given the role of  $f_2$ . Pairs and roles remain fixed for the length of the experiment.

For each pair, the program draws a card among Jack, Queen, King, Ace and Joker (all diamonds). This card is called the pair's card. It is common knowledge that the pair's card remains the same for the length of the experiment. In each pair,  $f_1$  and  $f_2$  are informed of the pair's card.

Across periods, each pair is matched to a different Audience and each pair is matched to each Audience exactly once. To avoid learning, no feedback about the outcome of the game is given between periods.<sup>1</sup> At the end of the session, one period is randomly selected that determines the payoff of every participant

---

<sup>1</sup>This is especially important for our analysis of the behavior of the Audience, which will mainly relies on within subject tests.

in the session. Participants are informed of their payment for the selected period only, and they are not informed of the payment of any other player. This insures that, within a pair, one participant cannot infer the decision of the other from his payment.

At the beginning of each period, the program draws a winning card for each pair independently, among Jack, Queen, King, Ace or Joker. This card can be different from the pair's card.  $f_1$  and  $f_2$  are informed of the winning card on their screen. Then,  $f_1$  and  $f_2$  are instructed to report their pair's card simultaneously and without any communication allowed. However, they can chose to report any card they want among Jack, Queen, King, Ace or Joker. Not reporting the card of the pair is cheating since it violates the rule. Participants are informed that their report is private, except for the Audience with whom they are matched for the present period.

After  $f_1$  and  $f_2$  have made their report, the Audience observes the card of the pair and the report of  $f_1$ , but not the report of  $f_2$ . Thus, the Audience is able to see whether  $f_1$  misreported, but not if  $f_2$  misreported. The Audience then decides whether to "pass" or "play".

$f_1$ 's payoff is the sum of two elements. If  $f_1$  reports the winning card, he earns 1125 ECU (125 ECU = €1). If  $f_1$  reports another card, he earns 875 ECU. The second element depends on the decision of the Audience: if the Audience decides to play,  $f_1$  earns an extra 625 ECU.

$f_2$ 's payoff depends on the decision of the Audience:  $f_2$ 's decision matters if and only if the Audience plays. If the Audience does not play,  $f_2$  earns a fixed amount of 750 ECU. If the Audience plays,  $f_2$  earns 1375 ECU when reporting the winning card, and 1000 ECU when reporting another card. Thus,  $f_1$  and  $f_2$  are always better off when the Audience plays.

If the Audience passes, his payoff is 875 ECU. If the Audience plays, his payoff depends on the report of  $f_2$ : if  $f_2$  reported truthfully, the Audience earns 1250



ECU. If  $f_2$  misreported, the Audience earns 625 ECU.<sup>2</sup>

## 2.2 Experimental treatments

Between our two conditions, we vary the potential for image spillovers. The only difference between the baseline and the treatment is that the members of each pair in the treatment are friends in the real life. This is not the case in the baseline.

In practice, we recruited couple of friends for both conditions. In the invitation email, it was made clear that participants should come to the lab with a friend, otherwise they would not be allowed to participate. In the treatment, the matching protocol was such that two friends were either both Audience or both members of the same pair. In the baseline, the matching protocol insured that two friends could never interact: two friends would either be both Audience or members of two different pairs. Participants were informed about the matching protocol in the instructions (see Appendix C).

## 2.3 Post experimental questionnaire

At the end of the session, but before any information about payoffs, participants had to make five guesses about how participants behaved in the session. In each of the five scenarios, there were two possible decisions. Participants were asked to predict the proportion of each decisions in their session. For instance, they had to predict the proportion of truthful reporting (resp. misreporting) by  $f_2$ s. Answers were incentivized: at the end of the session, one prediction was randomly drawn. The closer this prediction was from the true proportion, the more money they could earn.<sup>3</sup>

---

<sup>2</sup>In addition, after the fourth period of the game, participants played a mind coin toss game. The program simulated a random draw and participants had to report whether they predicted the outcome correctly. The aim was to test whether, in a pair of friends, the behavior of one player is correlated with the behavior of the other player in a context in which payoffs are independent. Since only 16,5% of the participants reported that they did not predict the outcome correctly, we do not use these data in the remaining of the paper. The data is available upon request.

<sup>3</sup>The full payoff function is as follows : If the prediction was equal to the actual value more or less 5 percentage point, they earned 250 ECU. If the prediction was between 5 and 10 percentage point away from the actual value, they earned 187.5 ECU. If the prediction was between 10 and

The first scenario corresponds to the decision made by  $f_1$ : participants had to predict the proportion of misreporting by  $f_1$ s. The next two scenarios correspond to the decision made by the Audience. Participants had to guess the proportion of decisions to play conditional on the report by  $f_1$ . First, they had to guess the proportion of decision to play after observing truthful reporting. Then, they had to guess the proportion of decisions to play after observing misreporting. The last two scenarios correspond to the decision made by  $f_2$ . First, they had to guess the proportion of decisions to misreport knowing that  $f_1$  misreported. Second, they had to guess the proportion of decisions to misreport knowing that  $f_1$  reported truthfully. If a participant expects that the decisions of  $f_1$  and  $f_2$  are independent, the two guesses about  $f_2$  should not differ. Conversely, if participants expect that for some reasons these decisions are not independent, the predictions should differ.

Then, participants had to fill out a questionnaire aimed at checking that friends who came together to the sessions were actually friends. To assess the degree of friendship of participant, we asked the same question to each participant about himself and about his friend (for instance “Do(es) you (your friend) have a pet?”) and compared answers to measure mutual knowledge. We also used a validated questionnaire by Mendelson and Aboud (1999) that measures the quality of friendship relations. Data from these questionnaires confirms that participants who came together were likely real-life friends (see appendix A for the details of these questionnaires and data supporting our claim).

Finally, we elicited socio-demographic characteristics, including a measure of risk attitudes using the procedure of Dohmen et al. (2011).

## 2.4 Experimental Procedures

The experiment was developed using z-Tree (Fischbacher, 2007). All sessions were conducted at GATE-LAB, Lyon. There were 10 sessions in total (4 sessions for the baseline, 6 for the treatment). For each session, we recruited 6 couples of friends.

---

15 percentage point from the actual value, they earned 125 ECU. If the prediction was between 15 and 20 percentage point from the actual value, they earned 62 ECU. If the prediction was further than 20 percentage points from the actual value, they earned nothing.

In total, 120 participants took part in the experiment. Participants were recruited using Hroot (Bock et al., 2014). In the recruitment email, we explained to the potential participants that they had to come with a friend to the laboratory in order to be allowed to take part in the session.

Participants were randomly assigned to individual cubicles. Friends were never assigned to adjacent cubicles. The instructions were distributed to participants and read aloud by the same experimenter (see Appendix C).

Before the first period, participants had to fill out a comprehension questionnaire. They could not proceed to the experiment before completing the questionnaire correctly. Questions were answered in private.

The average length of a session was 50 minutes. The average payoff was €14.78 (Min: €10, Max: €22, standard deviation, S.D. hereafter: 2.85). Payments were made in cash, in a separate room and in private.

### 3 Framework and hypotheses

To formalize the effect of the treatment on misreporting, we take the point of view of a participant  $i$  in the role of  $f_1$  who has to decide whether to misreport or report truthfully. Equation 1 represents his expected utility.

$$\mathbb{E}u_i(x, c, \theta_i) = \pi(x) - c(x, \theta_i) + p(x, c)M(c) \quad (1)$$

$x$  corresponds to his decision,  $x = m$  if he misreports and  $x = h$  if he reports truthfully.  $c$  represents the experimental condition, ( $c = b$  for the baseline and  $c = t$  for the treatment).

$\pi(x)$  represents monetary motivations. Participants can earn more money by misreporting if the winning card differs from the pair's card:  $\pi(m) > \pi(h)$ .  $c(x, \theta_i)$  is the moral cost of misreporting. We set  $c(h, \cdot) = 0$  and consider that  $\partial c(m, \theta_i) / \partial \theta_i > 0$ . This gives a direct interpretation to  $\theta_i$ , as an idiosyncratic preference parameter for truthful reporting. It captures the fact that participants

are heterogeneous in this respect.<sup>4</sup>  $c(x, \theta_i)$  also accounts for the potential affective value of having a good image in the eyes of other participants and the experimenter.

$p(x, c)M(c)$  formalizes the mechanism of image spillovers.  $p(x, c)$  represents the participant's beliefs on the decision of the Audience conditional on his report  $x$  and the experimental condition  $c$ . We define  $\Delta(c) = p(h, c) - p(m, c)$  as the impact of moving from misreporting to truthful reporting in condition  $c$ , as expected by  $f_1$ . We consider that  $\Delta(c) \geq 0$ . This means that the Audience is expected to be at least as likely to play after observing truthful report than after observing misreporting. We further assume that  $\Delta(t) > \Delta(b)$ . This captures the fact that participants expect the Audience to condition his decision on the decision of  $f_1$  to a greater extent in the treatment than in the baseline. We should stress that it does not imply that the Audience does not condition his decision in the baseline.<sup>5</sup>

$M(c)$  represents the fact that, in both experimental conditions, the participant prefers the Audience to play:  $M(c) > 0$ .  $M(c)$  represents the instrumental value for  $i$  of the image that the Audience forms about  $f_2$ : Audience plays when he thinks that it is likely enough that  $f_2$  reported truthfully and, in our design,  $f_1$  earns more money when the Audience plays. We consider that  $M(t) \geq M(b)$ :  $f_1$  is more concerned about the payoff of his friends than about the payoff of another anonymous participant.

$f_1$  is indifferent between misreporting and reporting truthfully if his type  $\hat{\theta}$  is such that equation 2 holds.

$$\mathbb{E}u_i(m, c, \hat{\theta}) = \mathbb{E}u_i(t, c, \hat{\theta}) \quad (2)$$

It is straightforward to show that for  $\theta_i < \hat{\theta}$ ,  $f_1$  misreports and for  $\theta_i > \hat{\theta}$ ,  $f_1$  reports truthfully. Equation 2 can be rearranged to give equation 3.

---

<sup>4</sup>For instance, Gibson et al. (2013) find an important heterogeneity in preference for truthfulness.

<sup>5</sup>There are several reasons that could lead Audience to condition his decision in the baseline. For instance, observing misreporting could prime the Audience with the idea that misreporting is a widespread behavior.

$$c(x, \hat{\theta}(c)) = \pi(m) - \pi(t) - \Delta(c)M(c) \quad (3)$$

$\hat{\theta}(c)$  is the type that makes participants indifferent between misreporting and reporting truthfully in condition  $c$ . Our comparative static exercise consists in comparing  $\hat{\theta}(b)$  and  $\hat{\theta}(t)$ . Equation 4 summarizes how the limit type is expected to change between experimental conditions.

$$c(x, \hat{\theta}(b)) - c(x, \hat{\theta}(t)) = \Delta(t)M(t) - \Delta(b)M(b) \quad (4)$$

Equation 4 implies that  $\hat{\theta}(b) > \hat{\theta}(t)$ : because image spillovers can occur in the treatment, some individuals who would misreport in the baseline report truthfully in the treatment. This is because his decision is expected to have a greater impact on the Audience’s decision in the treatment compared to the baseline ( $\Delta(t) > \Delta(b)$ ), and because the gain if Audience plays is at least as large in the treatment compared to the baseline ( $M(t) \geq M(b)$ ). This formalizes the (expected) deterrent effect of image spillovers.

From the previous framework, we derive three hypotheses.

**Hypothesis 1: The behavior of  $f_1$  is predictive of the behavior of  $f_2$  in the treatment, not in the baseline.** This hypothesis corresponds to testing whether friends are “similar”. Previous literature indicates that friends cluster by preference because of social influence and homophily (e.g. McPherson et al., 2001; Leider et al., 2009; Goeree et al., 2010), which should translate into behavior regarding rule breaking.

**Hypothesis 2: In the treatment, the Audience plays less frequently after observing misreporting than after observing truthful reporting. No such difference is expected in the baseline.** This hypothesis illustrates the fact that observing bad behavior on the behalf of a member of a social group harms the opportunity of other group members. We further conjecture that the effect is driven by *beliefs updating*: when  $f_1$  and  $f_2$  are friends, the Audience updates her beliefs about  $f_2$  upon observing how  $f_1$  behaved. An alternative mechanism would be that the Audience wants to punish individuals who misreported, in the spirit

of third party punishment, but by design, we made third party punishment very costly for the Audience and thus not likely to be a driver of behavior. Moreover, this motivation is in principle constant between experimental conditions.

**Hypothesis 3: There is less misreporting by  $f_1$  in the treatment compared to the baseline.** This hypothesis corresponds to the deterrent effect of image spillovers. Consistently with the framework presented above, we expect that some participants who would misreport in the baseline decide to report truthfully in the treatment. We expect them to internalize the harm their behavior can have on  $f_2$ 's image in the treatment.

## 4 Results

Table 1 summarizes the main variable of interest, namely the frequency of misreporting by  $f_1$  and  $f_2$ , and the decision to play by the Audience.<sup>6</sup>

Table 1: Summary statistics.

| Condition<br>Variable | Baseline |           |    | Treatment |           |    | Pooled |           |     |
|-----------------------|----------|-----------|----|-----------|-----------|----|--------|-----------|-----|
|                       | Mean     | Std. Dev. | N  | Mean      | Std. Dev. | N  | Mean   | Std. Dev. | N   |
| $f_1$ misreports      | 0.615    | 0.491     | 52 | 0.584     | 0.496     | 77 | 0.596  | 0.492     | 129 |
| $f_2$ misreports      | 0.673    | 0.473     | 52 | 0.649     | 0.480     | 77 | 0.658  | 0.475     | 129 |
| Audience plays        | 0.403    | 0.495     | 52 | 0.441     | 0.499     | 77 | 0.42   | 0.496     | 129 |

$f_1$  ( $f_2$ ) misreport=1 if  $f_1$  ( $f_2$ ) reported a card that was not his pair's card

Audience plays=1 if the Audience plays

**Result 1: The behavior of pair members is correlated only when pairs are composed of two friends.** This result supports Hypothesis 1.

**Support for result 1:** Figure 1 represents the proportion of misreporting by  $f_2$  for each report of  $f_1$  by condition. It suggests that the behavior of  $f_1$  and  $f_2$  are

<sup>6</sup>Unless otherwise stated, we drop observations in which subjects could not increase their payoff by misreporting, i.e. observations for which the winning card is the same as the pair's card: in this situation,  $f_1$  and  $f_2$  are always better-off reporting truthfully, and Audiences are always better-off playing. In these cases, participants played the dominant strategy 84.6% of the times. 9% of the participants deviated from the dominant strategies at least once (10.4% in the baseline, 8.3% in the treatment). Excluding them from our analysis does not change our results.

correlated in the treatment exclusively. In the baseline, the frequency of misreport by  $f_2$  goes from 60% when  $f_1$  reported truthfully to 71.9% when  $f_1$  misreported (Fischer exact test:  $p = 0.544$ ).<sup>7</sup> In the treatment, it goes from 46.9% when  $f_1$  reported truthfully to 77,6% when  $f_1$  misreported (Fischer exact test:  $p = 0.008$ ).

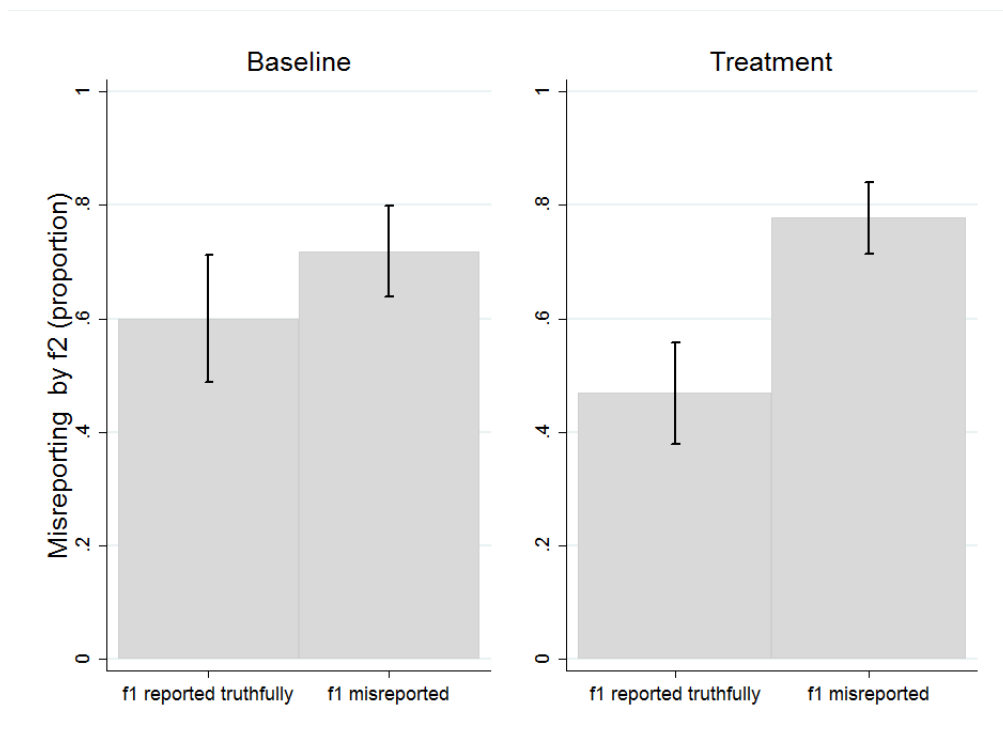


Figure 1: Proportion of misreporting by  $f_2$  conditional on  $f_1$ 's report, over condition.

Note: The vertical lines represent standard errors.

In order to account for multiple observations, we test whether decisions by  $f_1$  and  $f_2$  are more likely to match than if they were totally independent. First, we compute the theoretical (benchmark) proportion of matched decision for each pairs. We sum the expected proportion of the event “both  $f_1$  and  $f_2$  misreported” and “both  $f_1$  and  $f_2$  reported truthfully”. In each case, the expected proportion is computed

<sup>7</sup>A power analysis reveals that, for a power of 80%, in a one-sided fisher-exact test, we would need 446 observations in the baseline in order to have a significant effect at the 5% level.

as the product of the proportion of misreport (resp. truthful report) by  $f_1$  with the proportion of misreport (resp. truthful report) by  $f_2$ . Second, we compute the empirical frequency of matched decisions for each pair of  $f_1$  and  $f_2$ . We sum the empirical frequency of the event “both  $f_1$  and  $f_2$  misreported” and “both  $f_1$  and  $f_2$  reported truthfully”. Then, for each pair, we compare the empirical frequency to the theoretical proportion using a Wilcoxon matched-pairs signed-ranks test (W hereafter). Under the null hypothesis, the empirical frequency is not different from the theoretical one. Under the alternative, the empirical frequency is greater than the theoretical one.<sup>8</sup> Results are reported in Table 2.

Table 2: Comparison of empirical and theoretical frequencies of match between  $f_1$  and  $f_2$  decisions

|           | (1)                 | (2)                   | (3)   | (4) | (5)                  |
|-----------|---------------------|-----------------------|-------|-----|----------------------|
|           | Empirical frequency | Theoretical frequency | Diff  | N   | $p$ -value (1-sided) |
| Baseline  | 0.614               | 0.539                 | 0.075 | 16  | 0.189                |
| Treatment | 0.667               | 0.525                 | 0.142 | 24  | 0.047                |
| Pooled    | 0.645               | 0.53                  | 0.115 | 40  | 0.028                |

Column (1) reports the empirical frequency of match between decisions of  $f_1$  and  $f_2$  (either both misreported or both reported truthfully). Column (2) reports the theoretical frequency, i.e. the frequency of matches expected if the decisions of  $f_1$  and  $f_2$  were independent. Column (3) reports the difference between column (1) and (2). Column (5) reports the  $p$ -values for one-sided Wilcoxon matched-pairs signed-ranks test that the empirical frequency is greater than the theoretical one.

Overall, the frequency of matching decisions is greater than what is expected under the null (difference= 0.115,  $p = 0.028$ ). In the treatment,  $f_1$ 's and  $f_2$ 's decision match significantly more often than what is expected under the null (difference= 0.142,  $p = 0.047$ ). In the baseline, the difference between the empirical frequency and the theoretical frequency is not statistically significant (difference= 0.075,  $p = 0.189$ ).

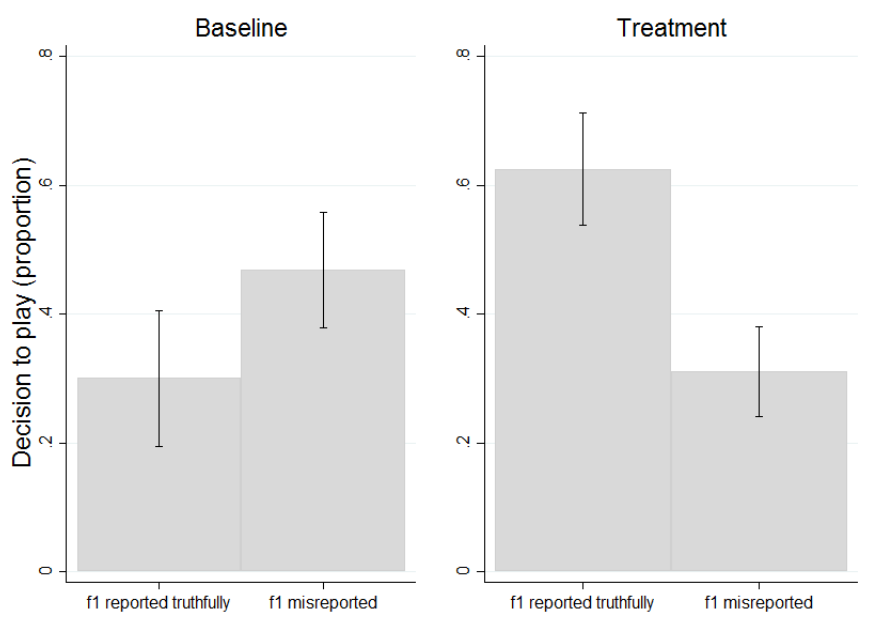
**Result 2: In the treatment, the Audience plays less often when  $f_1$  misreports. This is not the case in the baseline.** This result supports Hypothesis 2.

<sup>8</sup>Our test is one sided because our hypothesis follows the bulk of the literature on friendship that emphasizes that friends tend to be alike, rather than different.



**Support for result 2:** Figure 2 depicts the raw figures of decisions to play, conditional on the report by  $f_1$ , by condition.

Figure 2: Play decision conditional on report by  $f_1$ s, by condition



Note: The vertical lines represent standard errors.

In the baseline, Audience played 30% of the time after observing a truthful report and 46.9% after observing misreporting. In the treatment, Audience played 62.5% of the time after observing a truthful report and 31.1% after observing a untruthful report.

In order to formally test that the Audiences play less often when  $f_1$  misreports than when  $f_1$  reported truthfully, we compute for each Audience the frequency of decisions to play after observing misreporting and the frequency of decisions to play after observing a truthful report. We compare these frequencies within Audience. Under the null hypothesis, the frequency of decisions to play after observing misreporting is not different from the frequency to play after observing a truthful report.

The difference between the frequency of decisions to play after observing truth-

ful report and after observing misreport is not significant in the baseline (W:  $p = 0.618$ )<sup>9</sup> but is significant in the treatment (W:  $p = 0.016$ ). This suggests that Audiences are conscious of the correlation between the behavior of  $f_1$  and  $f_2$ , when  $f_1$  and  $f_2$  are friends. Table B.1 in Appendix B.1 confirms with logit regressions that this result is robust to the introduction of individual characteristics and session fixed-effects.

The theoretical motivation for the hypothesis supported by Result 2 is that the Audience can pull more information from the behavior of  $f_1$ s in the treatment than in the baseline. In order to explore this mechanism, we measure the extent to which beliefs about  $f_2$  is conditional on  $f_1$ 's report. For each Audience, we compute the difference between guesses when  $f_1$  reported truthfully and when  $f_2$  misreported. Audience guessed that  $f_2$ s were 4.75% less likely to misreport when  $f_1$  misreported in the baseline, but 9% more likely in the treatment. The previous comparison is qualitatively consistent with our expectations but does not reach significance (Two-sided Mann-Whitney test, MW hereafter,  $p = 0.13$ ).

A complementary analysis is to test if those who guessed correctly, i.e. those who, in the treatment, guessed that  $f_2$  would misreport more when the friend had misreported in the first place, are those who renounced to play after observing misreporting. To do so, we estimated logit models which explain the decision to play of the Audience by the Audience's guesses conditional on observing misreporting. The explanatory variables are a dummy variable indicating that  $f_1$  misreported interacted with the variable "update" that proxies the extent to which the Audience updates his beliefs about  $f_2$  after observing  $f_1$ . The variable "update" is the difference between the guess about the probability that  $f_2$  misreports after observing a truthful report and after observing a misreport. When the variable "update" is positive, the participants guessed that it is more likely that  $f_2$  will misreport after observing a misreport by  $f_1$  than after observing a truthful report. We also control for the belief of the Audience about the propensity of  $f_2$  to misreport upon observing a truthful report by  $f_1$ . In all specifications, we control for risk attitudes, gender, age, monthly expenses and whether participants have experience

---

<sup>9</sup>A power analysis reveals that, for a power of 80%, we would need 489 Audiences in the baseline in order to have a significant effect at the 5% level in a Mann-Whitney test.

with economic experiments. We also include session fixed effects. Table 3 reports the marginal effect of the variable “update” conditional on observing misreporting.

Overall, the variable “update” has a significant effect on the Audience’s decision to play conditional on  $f_1$  misreporting (column (1),  $p = 0.045$ ). Breaking down by experimental condition, the effect is not statistically significant in the baseline (column (2)), but is larger and highly statistically significant in the treatment (column (3)) ( $p < 0.001$ ). This suggests that the mechanism underlying result 2 relies on updating : Audiences pull more information from the observation of  $f_1$ , and use it more, when  $f_1$  and  $f_2$  are friends.

Table 3: The effect of updating on the decision to play conditional on observing misreporting by  $f_1$

|              | (1)                 | (2)               | (3)                   |
|--------------|---------------------|-------------------|-----------------------|
|              | <i>play</i> = 1     | <i>play</i> = 1   | <i>play</i> = 1       |
| update       | -0.004**<br>(0.002) | -0.003<br>(0.004) | -0.011****<br>(0.003) |
| Session F.E. | Yes                 | Yes               | Yes                   |
| Ind. char.   | Yes                 | Yes               | Yes                   |
| $N$          | 129                 | 52                | 77                    |
| Clusters     | 40                  | 16                | 24                    |

Marginal effect reported. Robust standard errors clustered at the individual level in parentheses.

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ , \*\*\*\*  $p < 0.001$

Since the behavior of  $f_1$  is predictive of the behavior of  $f_2$  in the treatment and the Audiences use this information, it is natural to test whether Audiences make “the right decision” more often in the treatment compared to the baseline. We consider that right decisions for the Audience are decisions that maximizes her payoff ex-post. There are two right decisions: (i) to play when  $f_2$  reported truthfully and (ii) no to play when  $f_2$  misreported. In order to assess whether Audiences make the right decision in the treatment more often , we computed the rate of right decisions for each individual Audience, and compared these rates between conditions. On average, Audiences make optimal decisions 48.43% of the times in the baseline and 61.15% of the times in the treatment (MW, two-sided  $p = 0.065$ ). This is the good news about image spillovers: external observers can pull valuable portable

information from the observation of the behavior of ingroups.

**Result 3: Image spillovers do not foster truthful reporting.** This result goes against hypothesis 3. Observed players do not adjust their behavior to the anticipation that their decision will influence the Audience’s decisions.

**Support for result 3:**  $f_1$  misreported 61,5% of the time in the baseline and 58,4% of the time in the treatment (misreporting rate are averaged at the individual level, Mann-Whitney test,  $p = 0.632$ ).<sup>10</sup> In Table B.2 in Appendix B, we present the result of logistic regressions in which misreporting by  $f_1$  is explained by the experimental condition. Results from these regressions confirm the absence of effect of image spillover on misreporting.

Result 3 sets off two potential explanations. The first one relates to preferences.  $f_1$ s could consider that it is not worthwhile to refrain from misreporting even if they expect that their behavior may impact the Audiences’ decisions: misreporting yields a certain payoff (250 ECU, €2) whereas truthful reporting has an uncertain impact on payoff, which depends on the Audience’s decision. This explanation suffers the following objection. If the Audience decides to play, it yields an amount of money to  $f_1$  that is considerably more important than misreporting (625 ECU for playing *vs.* 250 ECU for misreporting). In the treatment, more risk tolerant individuals should refrain from misreporting more often to have a chance to convince the Audience to play. We should thus see a relation between misreporting and risk tolerance. Spearman’s correlation coefficient between individual risk tolerance and individual misreporting frequency are very low and never significant if we pool all observations ( $\rho = -0.02$ ,  $p = 0.898$ ), or if we separate by condition (baseline :  $\rho = 0.088$ ,  $p = 0.745$ , treatment :  $\rho = -0.103$ ,  $p = 0.631$ ).

The second potential explanation relates to the perception of  $f_1$ s about the impact of their report on the decisions of the Audiences. As shown earlier, in the treatment, the Audiences played 61.5% of the time when observing truthful reporting and 31.1% of the times when observing misreporting, which represents a drop of

---

<sup>10</sup>A power analysis reveals that, for a power of 80%, we would need 1466 observations in total in order to have a significant effect at the 5% level in a Mann-Withney test.

31.4% percentage point.  $f_1$  guessed that Audiences would play 57% of the time when observing truthful reporting and 50.16% of the time after observing misreporting. This difference is significantly different from zero ( $W, p = 0.057$ ), but also significantly lower than the actual impact of observing misreporting on the decision to play ( $W, p < 0.001$ ).  $f_1$ s underestimate the impact of their report on the decision of the Audience. This in turn can explain why  $f_1$ s misreport as often in the treatment and in the baseline.

A possible source for these misspecified beliefs is that participants in the role  $f_1$  fail to realize that the behavior of  $f_1$ s predicts the behavior of  $f_2$ s in the treatment. Interestingly,  $f_1$ s are more likely than participants in other roles to guess that  $f_2$ s would report truthfully more often when  $f_1$  had misreported than when  $f_1$  had reported truthfully (Fischer exact test:  $p = 0.024$ ). An interpretation is that assignment to role  $f_1$  determined beliefs of participants in a self-serving manner, although belief elicitation was incentivized.<sup>11</sup> Believing that friends are actually different allows  $f_1$ s to hope that Audiences would play even after observing misreporting.

## 5 Conclusion

In this paper, we study how group members contribute to the image of the ingroup. In an experiment with pairs of friends, we tested whether individuals internalize the fact that they can harm their ingroup's opportunities by misbehaving if an observer infers a correlation between ingroups' decisions.

We find no evidence of such a mechanism. While external observers used the information contained in one's friend's behavior to update their beliefs about the other friend, participants did not internalize the impact of their behavior on their friend's image. This is the case although they have monetary incentives to preserve the image of the pair of friends they belong to. Our preferred interpretation is that participants strongly underestimate the harm they do to the economic opportuni-

---

<sup>11</sup>The impact of role assignment on beliefs has been studied before: Babcock et al. (1995) show in a dispute settlement experiment that what participants deem as fair depends on their role in the experiments.

ties of their friends. They expected their friend's economic opportunity to be only marginally affected by their behavior because they did not fully realize that they share a common image.

The results of the present paper contains a pessimistic message: even in situation in which a group member has substantial incentives to behave such as to preserve the group's image, mere ignorance of the mechanism leading to the production of group image can lead to hurt the group's image. As a consequence, the well identified disciplining effect of exposure found at the individual level (e.g. Andreoni and Bernheim, 2009; Bao et al., 2018; Kajackaite et al., 2018; Abeler et al., 2018) does not necessarily translate to situations in which image is collective rather than private.

In a more general perspective, our results identify a possible mechanism for poor social image at the group level. Group members fail to internalize that their behavior will impact the other members' opportunities. In our experiment, it is in their monetary interest to preserve the group's image so that this failure is not due to the fact that they do not care about their ingroups' fate, but to the fact that they fail to see that, whatever they do, it might be considered as a signal on the behavior of the members of their group.

Our results also contain a more positive message. In our experiment, strategical considerations related to the image of the group did not push dishonest participants to mimic honest participants. Thus, external observers were able to use the information contained in one's behavior to update their beliefs. Consequently, external observers took more profitable decisions in interacting with ingroups.

Several interesting related questions could be investigated. For instance, we could study image spillovers in the context of benevolent actions. Would external observers still use the information of one group member to update his beliefs about other group members? Would group members internalize the positive impact that their benevolent actions have on the image of the group?

## Acknowledgements

We are grateful to James Andreoni, John Hamman, Arno Riedl, Matthias Sutter, Adam Zylbersztejn, participants at the 2018 ESA world Meetings in Berlin, the 2018 ASFEE conference in Nice, the Cortex Student Club in Lyon, the BEER seminar at GATE for helpful discussions and to Marie Claire Villeval for detailed comments on earlier drafts. We thank Quentin Thevenet and Crissa Gresset for valuable support in running the experiment. This research was approved by the George Mason University IRB (ref: [1050112-1]). It has been supported by a grant of the French National Research Agency (ANR, FELIS Grant, ANR-14-CE28-0010-01), it was performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program Investissements d’Avenir (ANR-11-IDEX-007) operated by the French National Research Agency (ANR).

## References

- Abeler, J., D. Nosenzo, and C. Raymond (2018). Preferences for truth-telling. *Econometrica*.
- Aigner, D. J. and G. G. Cain (1977). Statistical theories of discrimination in labor markets. *ILR Review* 30(2), 175–187.
- Andreoni, J. and B. D. Bernheim (2009). Social image and the 50-50 norm: A theoretical and experimental analysis of audience effects. *Econometrica* 77(5), 1607–1636.
- Arrow, K. J. (1973). *The theory of discrimination*, pp. 3–33. Princeton University Press.
- Babcock, L., G. Loewenstein, S. Issacharoff, and C. Camerer (1995). Biased judgments of fairness in bargaining. *The American Economic Review* 85(5), 1337–1343.
- Bao, W., Y. Rao, J. Wang, and D. Houser (2018). Social exposure and trustworthiness: Experimental evidence. *Economics Letters* 162, 73 – 75.
- Bock, O., I. Baetge, and A. Nicklisch (2014). hroot: Hamburg registration and organization online tool. *European Economic Review* 71, 117 – 120.
- Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2016). Stereotypes. *The Quarterly Journal of Economics* 131(4), 1753–1794.
- Castillo, M. and R. Petrie (2010). Discrimination in the lab: Does information trump appearance? *Games and Economic Behavior* 68(1), 50 – 59.
- Christakis, N. A. and J. H. Fowler (2014). Friendship and natural selection. *Proceedings of the National Academy of Sciences* 111(Supplement 3), 10796–10801.
- Cohn, A., E. Fehr, and M. A. Maréchal (2014). Business culture and dishonesty in the banking industry. *Nature* 516(7529), 86.
- Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011). individual risk attitudes: measurement, determinants, and behavioral consequences. *Journal of the European Economic Association* 9(3), 522–550.



- Doleac, J. L. and L. C. Stein (2013). The visible hand: Race and online market outcomes. *The Economic Journal* 123(572), F469–F492.
- Eriksson, T., L. Mao, and M. C. Villeval (2017, Sep). Saving face and group identity. *Experimental Economics* 20(3), 622–647.
- Fershtman, C. and U. Gneezy (2001). Discrimination in a segmented society: An experimental approach. *The Quarterly Journal of Economics* 116(1), 351.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10(2), 171–178.
- Gibson, R., C. Tanner, and A. F. Wagner (2013, February). Preferences for truthfulness: Heterogeneity among and within individuals. *American Economic Review* 103(1), 532–48.
- Glover, D., A. Pallais, and W. Pariente (2017). Discrimination as a self-fulfilling prophecy: Evidence from french grocery stores. *The Quarterly Journal of Economics* 132(3), 1219–1260.
- Goeree, J. K., M. A. McConnell, T. Mitchell, T. Tromp, and L. Yariv (2010, February). The 1/d law of giving. *American Economic Journal: Microeconomics* 2(1), 183–203.
- Huck, S. and G. K. Lünser (2010). Group reputations: An experimental foray. *Journal of Economic Behavior & Organization* 73(2), 153–157.
- Kajackaite, A., U. Gneezy, and J. Sobel (2018). Lying aversion and the size of the lie. *The American Economic Review*.
- Kandel, D. B. (1978). Homophily, selection, and socialization in adolescent friendships. *American Journal of Sociology* 84(2), 427–436.
- Lee, K., M. C. Ashton, J. A. Pozzebon, B. A. Visser, J. S. Bourdage, and B. Ogunfowora (2009). Similarity and assumed similarity in personality reports of well-acquainted persons. *Journal of personality and social psychology* 96(2), 460.
- Leider, S., M. M. Möbius, T. Rosenblat, and Q.-A. Do (2009). Directed altruism and enforced reciprocity in social networks. *The Quarterly Journal of Economics* 124(4), 1815–1851.

- McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology* 27(1), 415–444.
- Mendelson, M. J. and F. E. Aboud (1999). Measuring friendship quality in late adolescents and young adults: McGill friendship questionnaires. *Canadian Journal of Behavioral Science* 31(2), 130–132.
- Phelps, E. S. (1972). The statistical theory of racism and sexism. *The american economic review* 62(4), 659–661.
- Tan, J. H. and C. Vogel (2008). Religion and trust: An experimental study. *Journal of Economic Psychology* 29(6), 832–848.
- Tirole, J. (1996). A theory of collective reputations (with applications to the persistence of corruption and to firm quality). *The Review of Economic Studies* 63(1), 1–22.

## A Assessment of friendship

In order to make sure that we were successful in recruiting friends to the lab, we collected two measures of friendship. The first one is original and relies on a measure of mutual knowledge. The second one is drawn from the literature in social psychology and measures the quality and the satisfaction of a given friendship.

### A.1 Measure of mutual knowledge

We designed a questionnaire that was aimed at measuring mutual knowledge. The underlying assumption is that the more two people (friends) interact, the better they know each other. It was implemented at the end of the session. Participants had to answer to a series of nine questions about themselves, and then they had to answer to the same questions about their friends. The questions are the following:

- The participant answers the following question about him
  1. Do you have a pet? (Yes, No)
  2. If you have a pet, what kind? (Open)
  3. If you have a pet, what is the name of the pet? (Open)
  4. Do you have any brother or sister? (Yes, No)
  5. If you have brother(s) and / or sister(s), how many? (Open)
  6. If you have brother(s) and / or sister(s), what is the name of the elder? (Open)
  7. Do you practice a sport? (Yes, No)
  8. If yes, what is this sport? (Open)
  9. If which season were you born? (Winter, Spring, Summer, Fall)
- The participant answers the following question about him
  1. Does your friend have a pet? (Yes, No, Don't know)
  2. If he/she has a pet, what kind? (Open)
  3. If he/she has a pet, what is the name of the pet? (Open)
  4. Does your friend have any brother or sister? (Yes, No, Don't know)

5. If your friend has brother(s) and / or sister(s), how many? (Open)
6. If your friend has brother(s) and / or sister(s), what is the name of the elder? (Open)
7. Does your friend practice a sport? (Yes, No, Don't know)
8. If yes, what is this sport? (Open)
9. If which season was your friend born? (Winter, Spring, Summer, Fall, Don't know)

By comparing the answers one gives about his friend and the friend's answers, we are able to compute how many times answers match. This proxies mutual knowledge. In case participants do not know each other, we expect answers to almost never match.

Under the null hypothesis that we recruited pairs of participants that were not actual friends, we can assume that they would answer randomly to the questions about their friends. In this case, we further assume that they would give correct answers to open questions ("What is the name of the pet?") with probability 0, and correct answers to closed questions with  $n$  possible answers with probability  $1/n$ . This assumption is conservative: because we let participants simply say that they don't know the answer to open questions, the probability that they give a good answer to such questions is probably lower than  $1/n$ . The previous assumption implies that, under the null hypothesis, we expect participants to give an average of 1.75 correct answers (S.D.: 5.062) and a maximum of 4 correct answers (with probability 0.031).

On average, participants gave 4.73 (S.D.: 2.33) correct answers about their friends, which is clearly greater than what we would expect under the null hypothesis (t-test,  $p < 0.001$ ). The former suggests that our recruitment procedure was successful in having actual friends coming to the lab.

## **A.2 Measure of the quality and satisfaction of a friendship**

We adapt the McGill questionnaire (Mendelson and Aboud, 1999). This test is aimed at measuring positive feelings of responder toward an identified friend. It

consists of item about which participant give their opinion on a 9-point likert scale. We selected the following 11 items from the original test:

1. You care a lot about your friend.
2. You like your friend a lot.
3. You want to stay friends for a long time.
4. You prefer your friend over most people you know.
5. You are happy you and your friend are friends.
6. You hope you will stay friends for a long time.
7. You would miss your friend if she/he left.
8. You enjoy him/her as a friend.
9. You are satisfied with your friendship.
10. You think your friendship is a great one.
11. You think your friendship is strong.

Opinion are between strongly disagree (-4) and strongly agree (+4). The scale has an excellent reliability (Cronbach's  $\alpha = 0.913$ ).

In the original study, casual friends had an average score of 2.2 and best friends had an average score of 3.5. In our study, participants had an average score of 2.976 (95% confidence interval : [2.77; 3.17]). This suggests that we were successful in recruiting casual friends at least.

## **B Supplementary analyses**

### **B.1 Marginal effect of observing misreporting on the decision to play.**

Table B.1 report the outcome of logit regressions for the marginal effect of observing misreport on the decision to play, over experimental condition. The first

column presents the simplest specification. In some specifications, we add individual characteristics (gender, risk attitudes, experience with experiment and monthly expenses) and session fixed-effects. The results presented in this table confirm Result 2: observing misreporting reduces the probability that the Audience plays in the treatment, but not in the baseline. Moreover, the marginal effect of observing misreporting is significantly greater in the treatment than in the baseline in every specification.

Table B.1: Effect of observing misreporting on the decision to play

|                                     | (1)                    | (2)                    | (3)                    | (4)                    |
|-------------------------------------|------------------------|------------------------|------------------------|------------------------|
|                                     | play=1                 | play=1                 | play=1                 | play=1                 |
| Baseline                            | 0.169<br>(0.197)       | 0.174<br>(0.196)       | 0.0953<br>(0.186)      | 0.166<br>(0.196)       |
| Treatment                           | -0.314****<br>(0.0927) | -0.324****<br>(0.0905) | -0.304****<br>(0.0858) | -0.360****<br>(0.0877) |
| <i>p</i> – value diff. <sup>a</sup> | 0.027                  | 0.021                  | 0.066                  | 0.017                  |
| Session F.E.                        | No                     | Yes                    | No                     | Yes                    |
| Ind. char.                          | No                     | No                     | Yes                    | Yes                    |
| <i>N</i>                            | 129                    | 129                    | 129                    | 129                    |
| Clusters                            | 40                     | 40                     | 40                     | 40                     |

Robust standard errors clustered at the individual level in parentheses

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ , \*\*\*\*  $p < 0.001$

## B.2 Treatment effect on misreporting

Table B.2 reports the outcome of logit regressions for the treatment effect on misreporting by  $f_1$ . The first column presents the simplest, univariate specification. In some specifications, we add individual characteristics (gender, risk attitudes, experience with experiment, and monthly expenses) and session fixed-effects. The results presented in this table confirms Result 3: there is no significant difference in misreporting by  $f_1$  between condition.

Table B.2: Misreporting by  $f_1$ s explained by treatment.

|              | (1)                | (2)               | (3)                | (4)               |
|--------------|--------------------|-------------------|--------------------|-------------------|
|              | $f_1$ mis.=1       | $f_1$ mis.=1      | $f_1$ mis.=1       | $f_1$ mis.=1      |
| Treatment    | -0.0310<br>(0.135) | -0.299<br>(0.254) | -0.0465<br>(0.115) | -0.374<br>(0.243) |
| Session F.E. | No                 | Yes               | No                 | Yes               |
| Ind. char.   | No                 | No                | Yes                | Yes               |
| $N$          | 129                | 129               | 129                | 129               |
| Clusters     | 40                 | 40                | 40                 | 40                |

Robust standard errors clustered at the individual level in parentheses

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ , \*\*\*\*  $p < 0.001$

## C Instructions for the treatment

Thank you for participating in this experiment on decision making. For the length of the session, please shut your phone down, and do not communicate unless we ask you to.

For showing up on time, you already get a 5 € show-up fee. This amount will be added to the payment you will get at the end of the session. During the experiment, you will have to take a number of decisions anonymously. Those decisions can make you earn money. Your earnings will be paid at the end of the session, in private in a separate room.

During the experiment, we will not talk about euros, but about experimental currency units (ECU). The exchange rate is :  $1\text{€} = 125\text{ ECU}$

The experiment is composed of three successive parts. This set of instructions is for the first part only. At the end of each part, you will be given the instructions for the next part.

Should you have any question, please use the call button situated in your cubicle. An experimenter will come to answer to your questions in private.

### First Part

The first part is composed of 4 periods.

At the beginning of the first period, the couple of friends who came together at the session will be divided in two. For half of the couple of friends, the program will randomly assign each member of the couple either the role of Friend 1 or the role of Friend 2. A Friend 1 and a Friend 2 constitute a pair of friends. Each pair of friends is composed of two friends who came together to the session. Each pair remains fixed for the duration of the first part: for each period, Friend 1 and Friend 2 came together to the session.

The second half of the couple of friends will be assigned the role of third party.



For each period, the interactions are between 3 participants: a pair of friends, composed of a friend 1 and a friend 2, and a third party.

For each period, each pair of friends interact with a different third party. Each pair of friends interacts exactly once with each third party and vice versa.

Before the first period, a playing card is drawn for each pair of friends among Jack of Diamonds, Queen of Diamonds, King of Diamonds, Ace of Diamonds, Joker. The drawing is independent for each pair of friends. The five cards are as likely to be drawn. The drawn card is associated to the pair of friends and this association remains fixed for the length of the first part, that is each pair of friends is associated to the same card for the 4 successive periods. The card associated to the pair is called the “card of the pair”. The five cards have exactly the same value.

Friends 1 and Friends 2 are informed, before the beginning of the first period, of their “card of the pair”, as shown in the following screen-shot:



Figure 1: Screenshots for the “card of the pair”

## Description of a period

At the beginning of each period, the software draws a playing card independently for each pair. This card is called “the winning card”. This card can be Jack of

Diamonds, Queen of Diamonds, King of Diamonds, Ace of Diamonds, Joker. Each card has the same probability to be drawn. The winning card can be the same from the pair's card, or it can be different.

The winning card is displayed on the screen of friends 1 and friends 2 of each pair, as shown in the following screenshot :

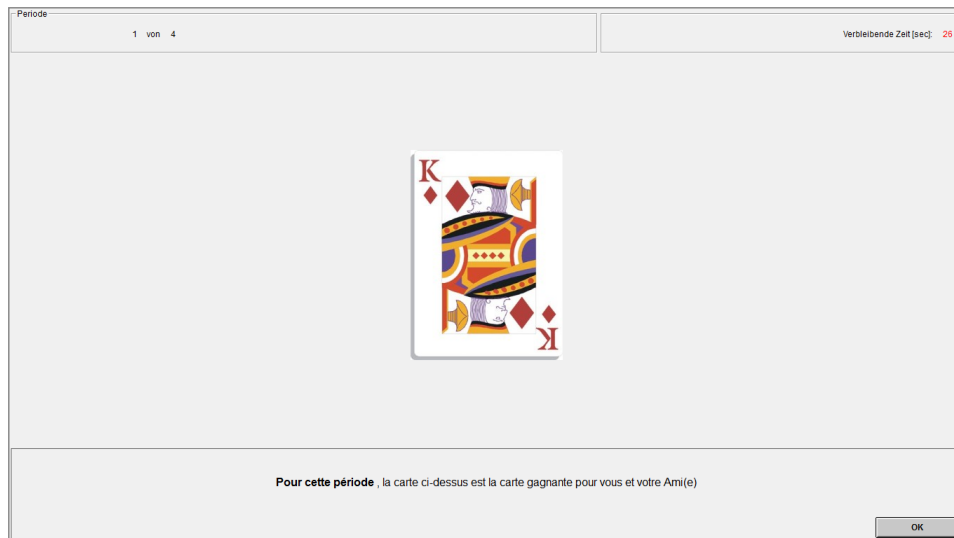


Figure 2: Screenshots for the “winning card”

After this, friend 1 and friend 2 must report the card of their pair simultaneously and without communicating, as shown in the following screenshot:

Once friend 1 and friend 2 in each pair have declared a card, the third party who is interacting with this pair for this period is informed of:

- The card of the pair.
- The winning card for this period.
- The card reported by friend 1.

The third party then must decide whether he wants to “implement” or not the report of friend 2. However, the third party cannot observe the report of friend 2. Implementing the report of friend 2 means that the report of friend 2 will determine the payoff of the third party and of friend 2 for this period. Not

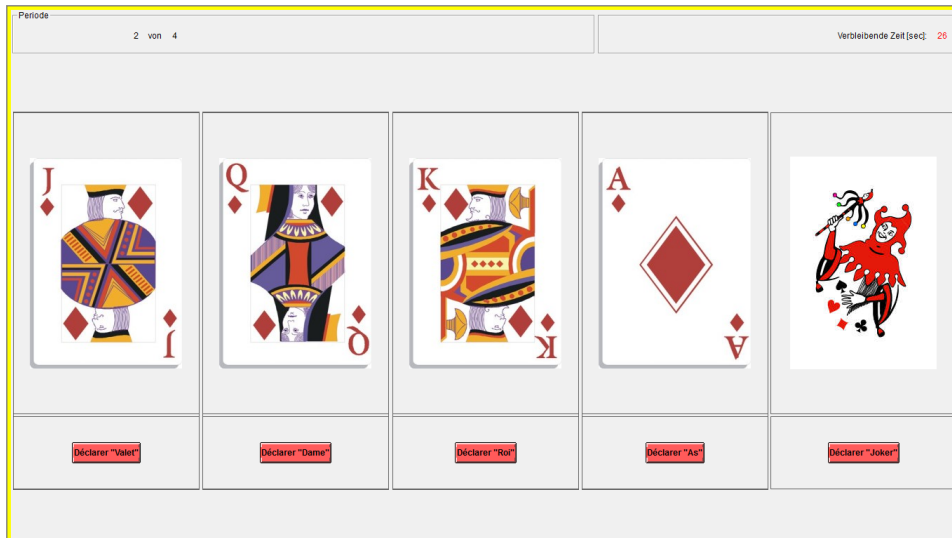


Figure 3: Screenshots for the “winning card”

implementing leads to another distribution of payoff, independent of the card of the par and of the winning card, as explained below. The following screenshots represent the decision screen for the third party:

The decision of the third party ends the period. No further information is revealed. A new period starts automatically. Participants will not be informed of their payoff before the end of the session, if this period is selected for payment.

### Determination of payoff

**Friend 1:** The payoff of friend 1 is composed of two elements:

- His report
  - If friend 1 declared a card that is not the winning card, he earns 875 UME.
  - If friend 1 declared the winning card, he earns 1125 UME.
- The decision of the third party:
  - If the third party implements the report of friend 2, friend 1 earns an extra 625 UME.

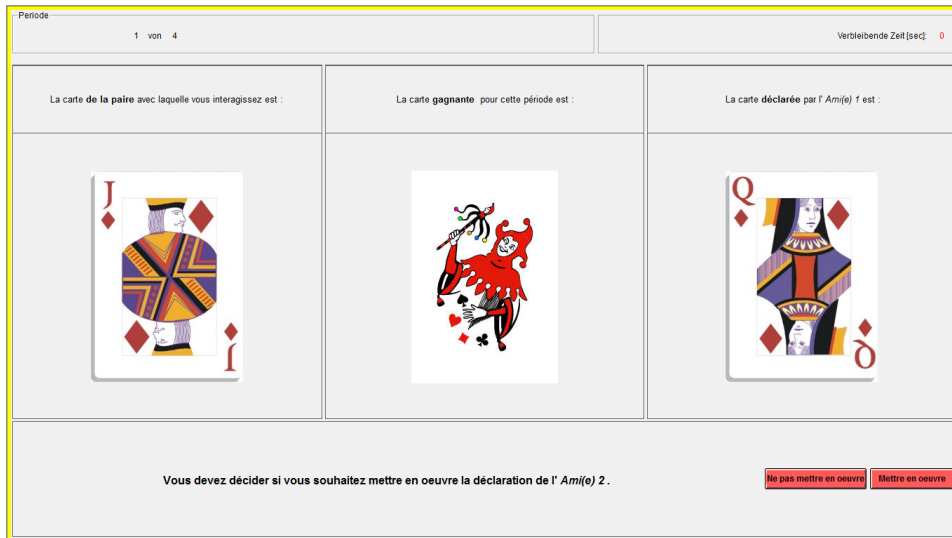


Figure 4: Screenshots: decision screen of the third party”

**Friend 2:** The payoff of friend 2 depends on the decision of the third party:

- If the third party decides not to implement the report of friend 2, friend 2 earns 750 UME.
- If the third party decides to implement the report of friend 2, the payoff of friend 2 depends on friend 2’s report:
  - If friend 2 reported the winning card, friend 2 earns 1375 UME.
  - If friend 2 reported another card, friend 2 earn 1000 UME.

**Third party** The payoff of the third party depends on his decision and, if the third party decides to implement, of the report of friend 2:

- If the third party decides not to implement, the third party earns 875 UME.
- If the third party decides to implement, then:
  - if friend 2 reported the card of the pair, the third party earns 1250 UME.
  - if fiend 2 reported another card, the third party earns 625 UME.

\*\*\*

Please read this instructions again. Should you have any question, please use the call button situated in your cubicle. And experimenter will come to answer your questions privately.

\*\*\*

## Part 2

In this part, you must predict the outcome of a coin toss. The coin is balanced: head and tail are equally likely.

In practice:

- You predict the outcome of the toss in your head.
- The software generate the toss and displays the outcome.
- You must indicate whether your prediction was correct.
- If your prediction was correct, you earn an extra 125 UME for this part.
- If your prediction was not correct, you eanr nothing for this part.

## Part 3

In this part, some situations from the experiment will be described on your screens. In each of these situations, participants had to choose one of two decisions.

In each case, you must predict the percentage of decisions by participants of the session corresponding to each option.

Below the description of the situation, you will see a slider. You must use this slider to input your predictions.

### Payoff for this part

At the end of the session, the software will randomly select one of your predictions. If this prediction is sufficiently close to the actual percentage, you earn money. The following rule is used to compute your earning for this part:

- If your prediction is equal to the true percentage  $\pm 5\%$ , you earn 250 ECU.
- If your prediction is equal to the true percentage,  $\pm 10\%$ , you earn 187.5 ECU.
- If your prediction is equal to the true percentage,  $\pm 15\%$ , you earn 125 ECU.
- If your prediction is equal to the true percentage,  $\pm 20\%$ , you earn 62 ECU.
- If your prediction is less accurate, you earn nothing for this part.

Example:

Participants had to choose between option A and option B. 65 % of decisions corresponded to option A (and 35 % to option B).

If you predict:

- 40% of decisions A, you earn nothing,

- 45% of decision A, you earn 62 ECU,
- 50% of decision A, you earn 125 ECU,
- 55% of decision A, you earn 187.5 ECU,
- 60% of decision A, you earn 250 ECU.

\*\*\*

Please read this instructions again. Should you have any question, please use the call button situated in your cubicle. And experimenter will come to answer your questions privately.

\*\*\*