Are groups always more dishonest than individuals? The case of salient negative externalities.

Geoffrey Castillo* Lawrence Choo† Veronika Grimm‡ §

11th May 2022

Abstract

A common finding of the literature on dishonesty is that groups are more dishonest than individuals. We revisit this finding by replacing the experimenter, implicitly hurt by subjects' dishonesty, with an explicit third-party: a local charity. With the charity we do not find groups to be more dishonest than individuals. Instead, groups can even help moderate the extent of the dishonesty.

JEL-classification: *C91, C92, D71.*
Keywords: Dishonesty, Group decisions, Communication, Social norms.

*Vienna Centre for Experimental Economics, Universität Wien, Austria.

†China Center for Behavioral Economics and Finance, Southwestern University of Finance and Economics, China. Corresponding author: `lawrence.cy.choo@gmail.com`

‡Department of Economics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany.

§We gratefully acknowledge the financial support of the Friedrich-Alexander-Universität Erlangen-Nürnberg through the Emerging Talent Initiative grant 2018/1_WW_01. We are grateful to the student assistants of the LERN laboratory who helped us to conduct the sessions. We are especially indebted to Michael Seebauer, without whom we could not have conducted the online sessions for the *Dictator* treatments. We thank Chew Soo Hong, Richard P. Ebstein, Todd R. Kaplan, Daniele Nosenzo, Andreas Roider, Silvio Staedter and Ro'i Zultan for comments on earlier versions. We also thank seminar participants at Southwestern University of Finance and Economics and Friedrich-Alexander-Universität Erlangen-Nürnberg, and conference participants for helpful comments.

1 Introduction

Studies of dishonesty in the laboratory often use the die-under-the-cup paradigm (Fischbacher and Föllmi-Heusi, 2013; see Abeler et al., 2019 for a review). This paradigm asks subjects to roll a die and report the number they see. Their payoff does not depend on the number seen but on the number reported, so subjects have an incentive to be dishonest. When groups instead of individuals complete the task, the literature finds that groups are more dishonest than individuals.[1]

In most of the studies using the die-under-the-cup paradigm to compare individuals and groups, dishonesty only harms the experimenter. Since this is not made explicit to the subjects, the negative externalities associated with dishonesty have little prominence. Outside the laboratory, however, those harmed by dishonesty are often explicit, for example when financial advisers mislead their clients or when aid workers steal from development funds. Negative externalities are thus more salient.

Salient negative externalities may lead to a different outcome in a group setting. Group members may abstain from expressing their preference for behaving dishonestly. Honest members may also more readily communicate moral reminders (Pruckner and Sausgruber, 2013). As a result, groups may become more honest than individuals.

We test this intuition by replacing the experimenter with a local charity. Since dishonesty now harms the charity there are clear real-world consequences and the negative externalities are more salient. We observe that, with the charity, groups are not more dishonest than individuals. In fact, groups can even help reduce the extent of the dishonesty. One of the reasons is that, with the charity, participants do not share opinions that support dishonesty.

We base our experimental design on Kocher et al. (2018). In this design subjects do not privately roll a die under a cup but see a video of a die roll. Kocher et al. (2018) observe similar levels of dishonesty compared to earlier studies where subjects roll their die privately. This design allows us to study not only the extensive margin—whether individuals or groups are dishonest—but also the intensive margin—how dishonest they are—for which we need to know the number subjects actually see. Some subjects decide on a number to report alone while others make their decision in a group: the between-subject comparisons enable us to study whether groups or individuals are more dishonest. Furthermore, group members

[1]See Barr and Michailidou (2017); Chytilová and Korbel (2014); Conrads et al. (2013); Gino et al. (2013); Kocher et al. (2018); Muehlheusser et al. (2015); Soraperra et al. (2017); Weisel and Shalvi (2015).

can communicate via chat to coordinate on a number to report, which allows us to study how communication shapes group dishonesty.

In our *Base* treatments, where dishonesty only harms the experimenter, we replicate the findings of the literature: while individuals and groups both report dishonestly, groups are much more dishonest than individuals.

We extend this design in our *Charity* treatments. Compared to the *Base* treatments, the monetary incentives remain the same but reporting a higher number now reduces the money donated to a local charity. Dishonesty is thus associated with clear negative externalities to the charity. We use a charity and not another subject because we want subjects to bring into the laboratory pre-existing "home-grown" norms (Schram and Charness, 2015) such as "one should not steal", which makes being dishonest unambiguously immoral.

Our first main result is that, with the charity, groups are not more dishonest than individuals. In fact, dishonest groups take significantly less from the charity than dishonest individuals. Our second main result is that dishonesty can be contagious: subjects who were members of a dishonest group are more likely to act dishonestly in the future.

To understand why the charity reduces dishonesty, we compare individual and group behaviour between the *Base* and *Charity* treatments. We find that the charity does not significantly affect individual dishonesty. This finding is consistent with Fischbacher and Föllmi-Heusi (2013) who observe that changing the consequences of dishonesty has little influence on individual dishonesty.

On the other hand, groups are more dishonest in the *Base* treatments than in the *Charity* treatments. Looking at the chat data, we find that subjects in the group *Charity* treatment send fewer messages calling for a dishonest report than those in the *Base* treatment. This suggests that the shift in dishonesty is driven by group members being more reluctant to appear dishonest when the charity is involved.

We explore other explanations in additional treatments. In the *CharityR* treatments we increase the negative externality even further by sending the charity an anonymised copy of the decisions made by the subjects and of the messages exchanged in the chat. We do not find any strong differences between the *Charity* and the *CharityR* treatments, suggesting that groups not being more dishonest than individuals comes mostly from the interactions in the group chat. It is also possible that the *CharityR* treatments suffer from low-intensity: subjects might not care about the charity being hurt by learning that they made dishonest decisions, and so the treatment might not increase the negative externality as intended. Introducing the charity might also have already made groups as honest as possible.

Another explanation is that the charity may make groups more pro-social than individuals. Previous research has indeed found that groups are more pro-social than individuals (Cason and Mui, 1997) and introducing the charity might have triggered other-regarding preferences in groups. In our design being pro-social means not over-reporting, which is confounded with not being dishonest. We test this alternative explanation in the *Dictator* treatments, where we remove the die roll and simply ask individuals and groups to split money between themselves and the charity. We do not find any differences between individuals and groups, which suggests that, in our setup, groups are not significantly more pro-social than individuals.

These findings relate to the literature on dishonesty that uses, not the die-rolling task, but the sender-receiver game of Gneezy (2005). In this game, the sender must send a message to the receiver and can lie to get a higher payoff, so by design there are consequences for a third-party and thus salient negative externalities. Similarly to what we observe, individuals and groups do not differ in terms of dishonesty (Sutter, 2009).

Our results point to the salience of negative externalities as an important determinant of dishonesty in groups. In some situations, negative externalities are not salient, for example when the victim of dishonesty is not easily identifiable. Group members can easily disregard the negative externalities and convince themselves they can behave dishonestly, even more so if moral wiggle room is present. In other situations, negative externalities are more salient. This happens when dishonesty hurts a clear victim, such as the charity in our *Charity* treatments. Here, it is much harder for group members to look the other way.

There exist two other papers that use a charity in the die-under-the-cup paradigm, but none involve groups. Maggian (2019) compares individual dishonesty with and without negative externality to a charity—essentially our individual *Base* and *Charity* treatments—and finds no difference, which is in line with our results. Drouvelis and Pearce (2021) find that people scoring high on an intelligence test are honest when lying benefits themselves, but dishonest when lying benefits a charity.

In the next Section we present the *Base*, *Charity* and *CharityR* treatments. Section 3 reports the results from these treatments. In Section 4 we consider pro-sociality as an explanation for our results and present the *Dictator* treatments. Section 5 reports the post-study probability to assess the reproducibility of our findings. Finally, Section 6 concludes.

Table 1: Payoffs to subjects and donations to the charity.

Die	⚅	⚀	⚁	⚂	⚃	⚄
Points	0	1	2	3	4	5
Payoff to subject	0 €	2 €	4 €	6 €	8 €	10 €
Donation to charity (*Charity* and *CharityR* treatments)	10 €	8 €	6 €	4 €	2 €	0 €

2 Experimental design

2.1 Treatments

Our main experiment involves six treatments, grouped into three categories: *Base*, *Charity* and *CharityR*. Figure 1, which we will use throughout, summarises our design.

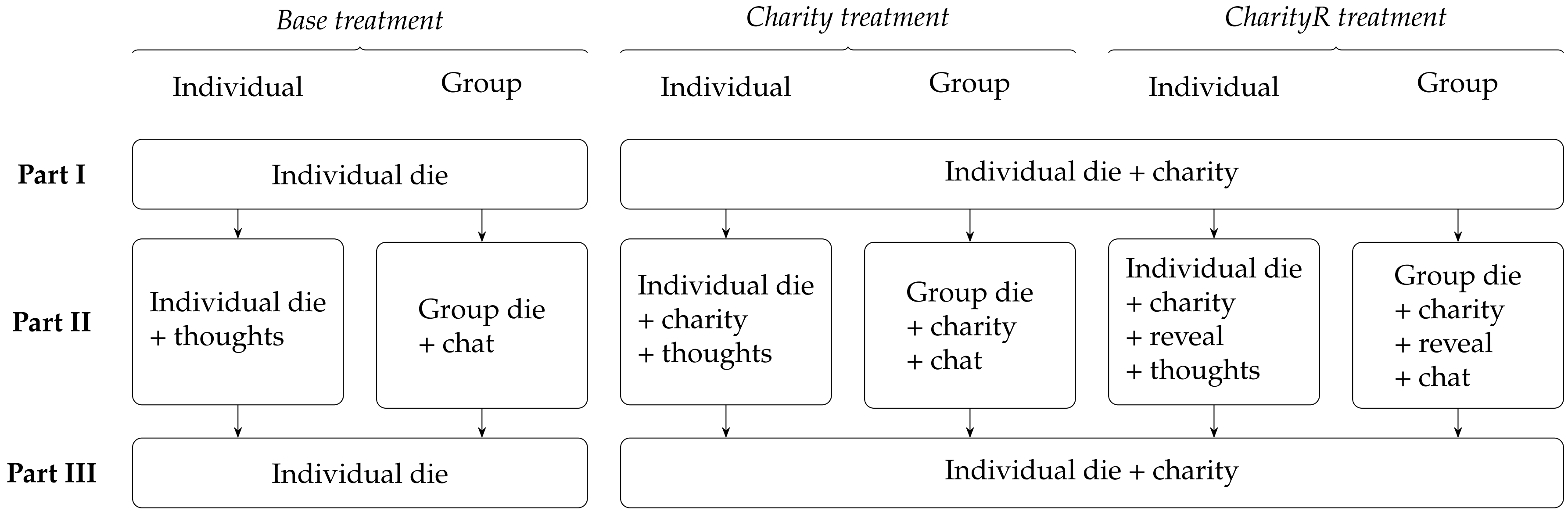
***Base* treatments.** To the left of the Figure are our *Base* individual (*I-Base*) and group (*G-Base*) treatments, which replicate the "IndividualDeliberation" and "GroupPC" treatments of Kocher et al. (2018).

There are three Parts to each treatment. In Parts I and III, subjects watch a video of a die roll and are asked to *truthfully* report the result of the roll which can be ⚀, ⚁, ⚂, ⚃, ⚄ or ⚅ with equal probability. Table 1 shows that the reported die-roll number is converted into points so that ⚅ corresponds to 0 points, ⚀ to 1 point, ⚁ to 2 points, and so on. Each point is worth 2 €.

What differs between the *I-Base* and *G-Base* treatments is the second part. In the *I-Base* treatment, subjects do the same individual task as in Parts I and III. In the *G-Base* treatment, however, subjects form groups of three, all observe the same die-roll video, and then independently report what they observed. Group members each receive the points shown in Table 1 if they all reported the same number, and 0 points otherwise. For example, each group member receives 3 points (6 €) if all three report ⚂. On the other hand, each group member receives 0 points (0 €) if they fail to report the same number.

To coordinate, group members can chat anonymously for up to 5 minutes through the experimental software. To keep the treatments comparable, we also follow Kocher et al. (2018) and give 5 minutes to subjects in Part II of the *I-Base* treatment to, if they wish, type their thoughts about the experiment.

Because subjects know that the experimenter knows which video they saw, this design could elicit a lower level of dishonesty. Note, however, that we focus,



Notes. **+ charity**: subjects' decisions also affect the donations the experimenter makes to a local charity.
+ thought: subjects can write their thoughts about their decision (for maximum 5 minutes).
+ chat: subjects of the same group can chat (for maximum 5 minutes).
+ reveal: subjects' decisions are revealed to the charity.

Figure 1: Summary of the experimental design.

not on absolute dishonesty, but rather on the difference in dishonesty between individuals and groups.

***Charity* treatments.** The *Charity* individual (*I-Charity*) and group *(G-Charity)* treatments are identical to their respective *Base* treatments, except they make explicit that reporting a larger number hurts a local charity: each extra Euro given to the subject as a result of their report is not given to a local charity. For example, as shown on Table 1, if a subject reports ⚂ the subject gets 6 € and the charity gets 4 €. In Part II of the *G-Charity* treatment, group members and the charity receive 0 € if the group members do not report the same die number.

***CharityR* treatments.** In the *CharityR* treatments, we further reveal to the charity the die number observed and reported in Part II of the individual (*I-CharityR*) and group (*G-CharityR*) treatments, as well as the written thoughts (*I-CharityR*) and chat logs (*G-CharityR*). All of this information is anonymous. Parts I and III are unchanged.

2.2 Procedures

When the experiment started, subjects were told that the experiment was made of three independent parts. At the beginning of each part they received specific instructions and saw a table similar to Table 1, except that points were not converted to Euros. To make *I-Charity* and *G-Charity*, and *I-CharityR* and *G-CharityR,* as comparable as possible, subjects faced the same table in both treatments. To ensure proper understanding of the instructions, subjects had to go through control questions under three different scenarios. For each of them, they had to answer how much they would get, how much the charity would receive from them, and, in *G-Charity* and *G-CharityR*, how much the charity would receive in total from the group. The control questions also made clear that, even if their stated task was to truthfully report the die roll, their payoffs, and the donations to the charity, only depended on their report.

We chose a small local charity whose members perform as clowns in nearby hospitals to entertain sick children. This choice minimises the chance that subjects lied to indirectly favour themselves, for instance if we had chosen a student organisation. In the initial instructions of the *Charity* and *CharityR* treatments, we gave subjects a broad description of the charity; they were then given 5 minutes to visit its website and Facebook page to learn about its activities. In the *CharityR* treatments, we also gave subjects a sample of the information that we would send to the charity via email. Subjects knew they would be included in blind carbon copy to the email. Since our intention was to be as close as possible to Kocher et al. (2018) in our baseline, we did not include information about the charity in the *Base* treatment.

We conducted the experiment at the LERN of the University of Erlangen-Nuremberg between June 2018 and March 2019. Subjects were recruited via ORSEE (Greiner, 2015). We recruited 30 subjects for each of the individual treatments and 90 subjects for each of the group treatments, ensuring the same number of independent observations in each case. We programmed the experiment with zTree (Fischbacher, 2007).[2] We present the instructions in Appendix E.2.

Following Kocher et al. (2018) we randomly generated in the first session one sequence of die rolls for each group and used the same sequence in the next sessions. We can therefore directly compare the die-roll reports across treatments. This procedure also ensured that subjects of the same group observed the same die rolls in Parts I and III.

At the end of each session, one of the three parts was selected at random to determine the earnings of the subjects and the donations to the charity. To convince subjects that the charity and the donations were real, we made the anonymous donations immediately from within the laboratory, and subjects were actively encouraged to monitor the process. Subjects knew this when they started the experiment.

Each session lasted approximately one hour. The mean earnings for subjects in the *Base*, *Charity* and *CharityR* treatments, including a show-up payment of 4 €, were 13.45 €, 12.06 € and 10.84 €, respectively.

2.3 Predictions

Since our *Base* treatments are based on Kocher et al. (2018), we expect that groups will report more points than observed compared to individuals. Kocher et al.'s (2018) favoured explanation is that "it is the exchange of arguments and moral views within the group that shift group members' expectations and behavior". The same might happen in the *Charity* and *CharityR* treatments, and so groups may across the board be more *dishonest* than individuals. In addition, and in particular with the charity, diffusion of responsibility and the bystander effect might play a role (Choo et al., 2019; Darley and Latàn, 1968). When individuals report a higher number than the one they have seen, it is clear that their own actions will result in

[2]We are grateful to Martin Kocher, Simeon Schudy and Lisa Spantig for making their zTree code available to us, which we adapted to our experiment.

the charity receiving less than it is due. But in a group setting, the bystander effect kicks in: even if they had wanted their group to report truthfully, they can instead choose to remain passive and acquiesce to the other members who want to report a higher number since the charity not receiving a lot of money is not only their fault but also the fault of the group.

On the other hand, with the charity, groups might become more *honest* than individuals. In the *G-Base* treatment, there is no reason for groups who observe less than 5 points to report something different that what they observed if they want to be honest, or 5 points if they want to maximise their earnings. In the *G-Charity* and *G-CharityR* treatments, some groups may still report more than what they observed but not all the way to 5 points to ensure that the charity receives something. Even if all group members prefer to report 5, not wanting to be seen as the one wanting to steal from a deserving cause can make it difficult to express such preference. The group's decision may thus be honest even when group members would have over-reported when making the decision alone.

3 Results

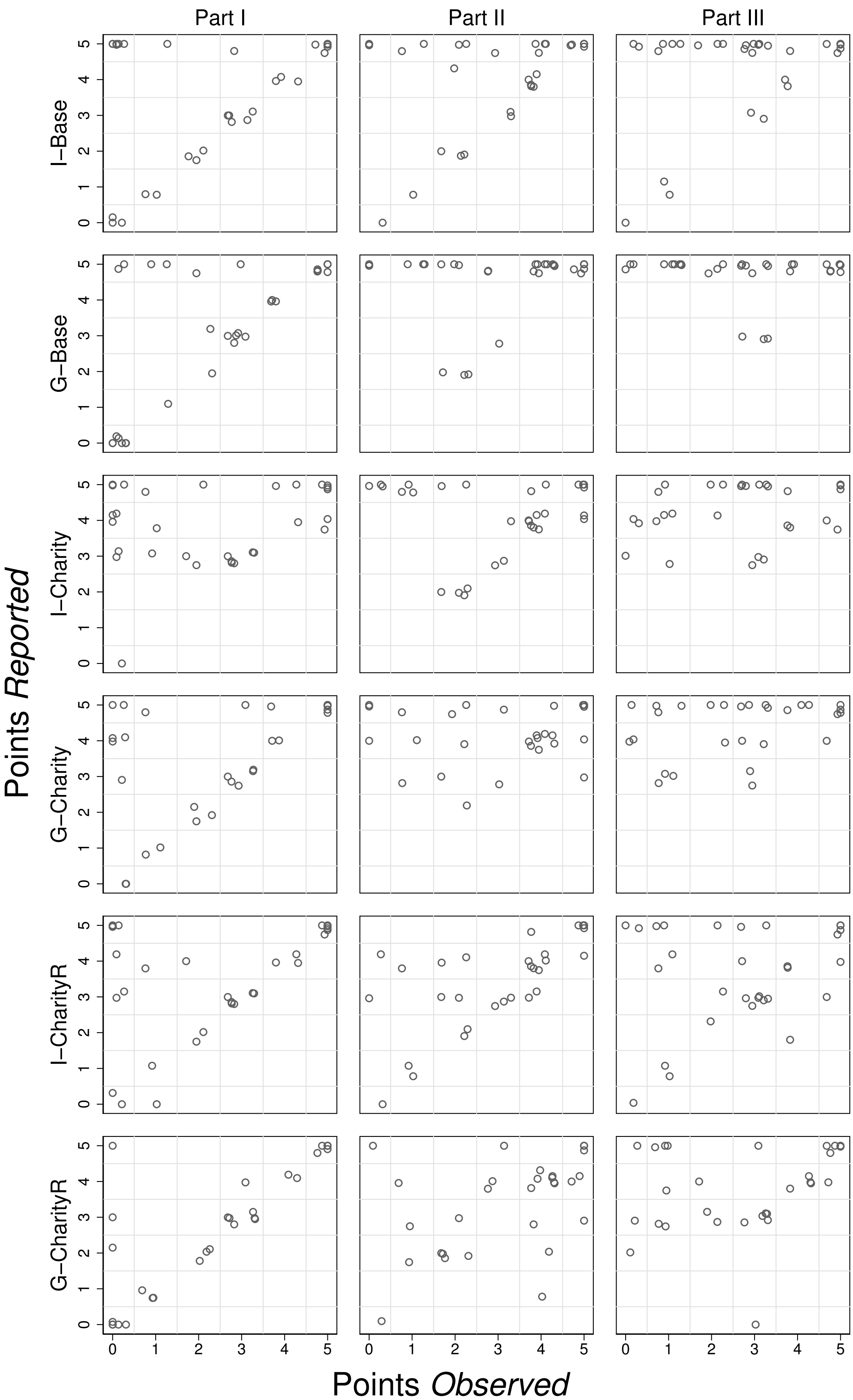
Figure 2 shows, for all treatments and all parts, the numbers reported (y-axis) as a function of the numbers seen (x-axis). In the group treatments and to compare with the individual treatments, we follow Kocher et al. (2018) and look at the median of the points reported in each group.

3.1 Groups are not always more dishonest than individuals

For the time being, we focus on Part II. Here almost every group managed to coordinate: 100%, 93% and 97% of the groups in the *Base*, *Charity* and *CharityR* treatments reported the same number. We thus omit observations in which members failed to coordinate in Part II.

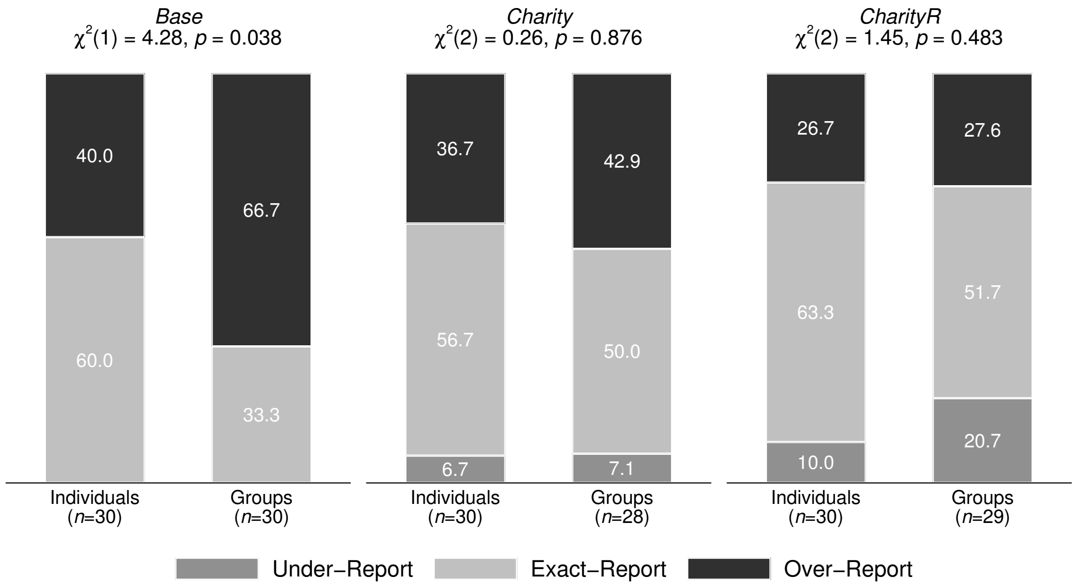
We categorise individuals and groups by comparing the number they reported to the one they saw. Individuals and groups *over-report* if their report gets them more points that if they had reported honestly. Similarly, they *under-report* if it gets them less, and they *exact-report* if they report the number seen. Figure 3 shows the proportion of each category in Part II. The Figure also reports the Pearson's test statistics associated with the comparison of the categories between individuals and groups in each treatment.

Since we are interested in the influence of negative externalities on dishonest



Note. In the individual treatments, each dot represents the point observed and reported by a participant. In the group treatments, each dot represent the point observed (all group members observe the same points in Parts I, II and III) and the median point reported by a group.

Figure 2: Points observed and points reported across treatments and parts (jittered).



Note. Pearson's $\chi^2(1)$ tests.

Figure 3: Proportion of over-, under- and exact-reporting individuals and groups in Part II.

behaviour, we will focus on over-reporting. Note that we observe no under-reporting in the *Base* treatments. We observe some under-reporting in the treatments involving the charity, especially in the *CharityR* treatments. There, under-reporting allows subjects to give to the charity more than if they had reported honestly. As indicated in the Figure, however, we do not find significant differences between the individual and group variants of the *Charity* and *CharityR* treatments even when taking into account under-reporting. There are also no differences between individuals and groups: the proportion of individuals and groups who under-report are 6.67% and 7.14% ($\chi^2(1)$ = 0.005, p = 0.943) in the *Charity* treatments, and 10.00% and 20.69% ($\chi^2(1)$ = 1.303, p = 0.254) in the *CharityR* treatments.[3]

As in Kocher et al. (2018), we find that in the *Base* treatments groups over-report more than individuals: 67% of groups over-report compared to 40% of individuals ($\chi^2(1)$ = 4.286, p = 0.038; all statistical tests are two-sided). Note that

[3]All subjects made individual decisions in Part I. Here, we do not find any significant differences in the proportion of individual and group treatment subjects who over-report in the *Base* (27% of individuals vs 29% of groups, $\chi^2(1)$ = 0.054, p = 0.815) and *CharityR* (30% of individuals vs 22% of groups, $\chi^2(1)$ = 0.742, p = 0.389) treatments. There is, however, some differences in the *Charity* treatments (52% of individuals vs 36% of groups, $\chi^2(1)$ = 2.963, p = 0.085), where individuals are slightly more dishonest than groups. In the Appendix we thus also report regressions controlling for behaviour in Part I.

our die sequences are different from the ones used by Kocher et al. (2018) (see Appendix E.1 for a comparison), which further strengthens the replicability of their findings:

Observation. *Groups over-report more frequently than individuals when over-reporting harms the experimenter.*

In the treatments involving the charity, however, we do not find that groups over-report more frequently than individuals. Figure 3 shows that the proportions of over-reporting groups and individuals are 43% and 37% ($\chi^2(1) = 0.231$, $p = 0.630$) in the *Charity* treatments, and 28% and 27% ($\chi^2(1) = 0.006$, $p = 0.937$) in the *CharityR* treatments. The differences remain insignificant ($\chi^2(1) \leqslant 0.348$, $p \geqslant 0.555$) even if we exclude instances where subjects observed a die number corresponding to 5 points, for which they could not have over-reported.

If we look, not at the proportion of individuals or groups who over-report, but at the reported points, a similar picture emerges: groups report significantly higher points than individuals in the *Base* treatments (Mann-Whitney $U = 324.5$, $p = 0.019$), but there are no significant differences in the reported points of individuals and groups in the *Charity* ($U = 407$, $p = 0.828$) and *CharityR* ($U = 419.5$, $p = 0.807$) treatments. These observations lead us to our first result:[4]

Result 1. *There is no significant difference in over-reporting between groups and individuals when over-reporting harms the charity. This holds even when the charity is informed.*

To give more meaning to these numbers, we can look at the extra revenue individuals and groups extract by over-reporting (Appendix A). In Part II of the *Base* treatments, individuals on average claim 2.14 € more than if they had reported honestly, and group members, 3.26 €. In line with the previous findings, the difference shrinks in the *Charity* treatments: individuals claim 2.26 € more, but group members now claim only 1.80 € more. In the *CharityR* treatments, the numbers for individuals and group members are even lower, respectively 0.94 € and 0.26 €.

It is clear from Figure 3 that individuals increasing their over-reporting with the charity does not drive the results: the proportion of over-reporting individuals

[4]The result does not change even if we consider a broader definition of dishonesty and categorise both over- and under-reporting as being dishonest: With this definition the proportion of dishonest reports is 43% in *I-Charity* versus 50% in *G-Charity* ($\chi^2(1) = 0.258$, $p = 0.611$); and 37% in *I-CharityR* versus 48% in *G-CharityR* ($\chi^2(1) = 0.81$, $p = 0.367$).

in the *Base* treatment is not significantly different from those in the *Charity* ($\chi^2(1) = 0.070$, $p = 0.791$) and *CharityR* ($\chi^2(1) = 1.200$, $p = 0.273$) treatments.[5]

Instead, Figure 3 shows that the proportion of over-reporting groups in Part II falls sharply when we introduce the charity: the proportion of over-reporting groups in the *Base* treatment is significantly higher than those in the *Charity* ($\chi^2(1) = 3.320$, $p = 0.068$) and *CharityR* ($\chi^2(1) = 9.032$, $p = 0.003$) treatments.

Also, revealing subjects' behaviour to the charity has little influence: the proportion of over-reporting individuals in Part II of the *I-Charity* and *I-CharityR* treatments are not significantly different ($\chi^2(1) = 0.693$, $p = 0.405$). Similarly, the proportion of over-reporting groups in Part II of the *G-Charity* and *G-CharityR* treatments are not significantly different ($\chi^2(1) = 1.458$, $p = 0.227$).

These observations lead us to the next result:

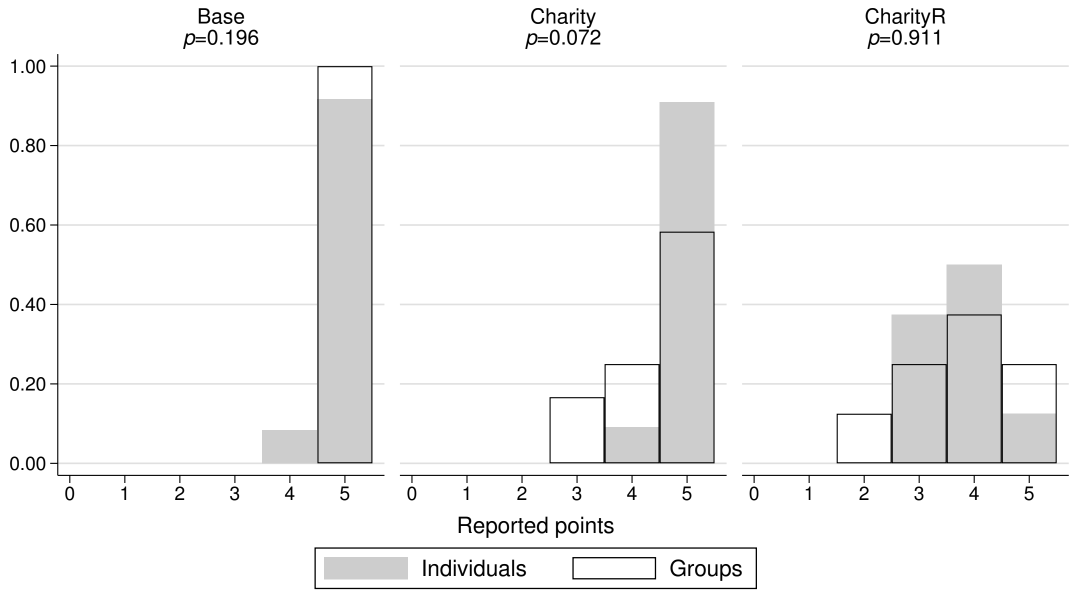
Result 2. *The proportion of over-reporting groups, but not of over-reporting individuals, decreases with the charity. Revealing anonymously to the charity the die number observed and reported, as well as the written thoughts and the chat logs, has no effect.*

We have established that there is no significant differences in the proportion of over-reporting groups and individuals—the extensive margin—when there are salient negative externalities. We now focus on the size of the lies—the intensive margin—amongst the over-reporting groups and individuals. There are no significant differences in the points they observed in the *Base* ($U = 106.5$, $p = 0.583$), *Charity* ($U = 64.5$, $p = 0.924$) and *CharityR* ($U = 29.5$, $p = 0.787$) treatments, so we can concentrate on the points they report.

Figure 4 shows the histogram of the reported points of these over-reporting individuals or groups. In the *Base* treatments, they report a similar number of points ($U = 110$, $p = 0.197$): almost always the maximum possible. This is also the case in Kocher et al. (2018). In the *Charity* treatments, however, over-reporting groups report significantly lower points ($U = 43.5$, $p = 0.072$) than over-reporting individuals. The difference remains significant ($U = 32$, $p = 0.089$) even after excluding instances where over-reporting individuals or groups observe 4 points, for which they could only report 5 points. In the *CharityR* treatments the effect is in the same direction but it is not significant ($U = 31$, $p = 0.911$). Hence our third result:

Result 3. *Over-reporting groups over-report less than over-reporting individuals when group members know that over-reporting harms the charity.*

[5]We observe the same in Part I: the proportion of over-reporting individuals in the *Base* treatments is 28%, compared to 32% for the *Charity* and *CharityR* treatments ($\chi^2(1) = 0.527$, $p = 0.468$), which we pool since at this point subjects faced the exact same task.



Note. p-values from Mann-Whitney U tests comparing the reported points of over-reporting individuals and groups.

Figure 4: Histogram of reported points by over-reporting individuals and groups in Part II.

In Appendix B.1.1, we report regressions corresponding to Results 1, 2 and 3, where we further control for the die roll observed in Part II, the behaviour in Part I, and gender. All the results carry through.

In Figure 4 we can also see that, when moving from the *Base* treatment to the *Charity* treatment, the points reported by over-reporting groups fall sharply ($U = 70$, $p = 0.002$) while those reported by over-reporting individuals remain stable ($U = 65.5$, $p = 0.949$). When moving from the *Charity* treatment to the *CharityR* treatment, however, it is now the points reported by over-reporting individuals that fall ($U = 8$, $p < 0.001$) while those reported by over-reporting groups are not significantly different ($U = 29.5$, $p = 0.127$).

3.2 Analysis of the chat data

We have established that groups do not over-report more often than individuals when we introduce a charity. To better understand why, we look at the chat data.

Two research assistants independently *coded* chat messages: they assigned a number to each chat message when subjects proposed a number or when they agreed with a previous proposal (see Appendix C for the full details). Between 32% and 33% of chat messages were coded in the *Base*, *Charity* and *CharityR*

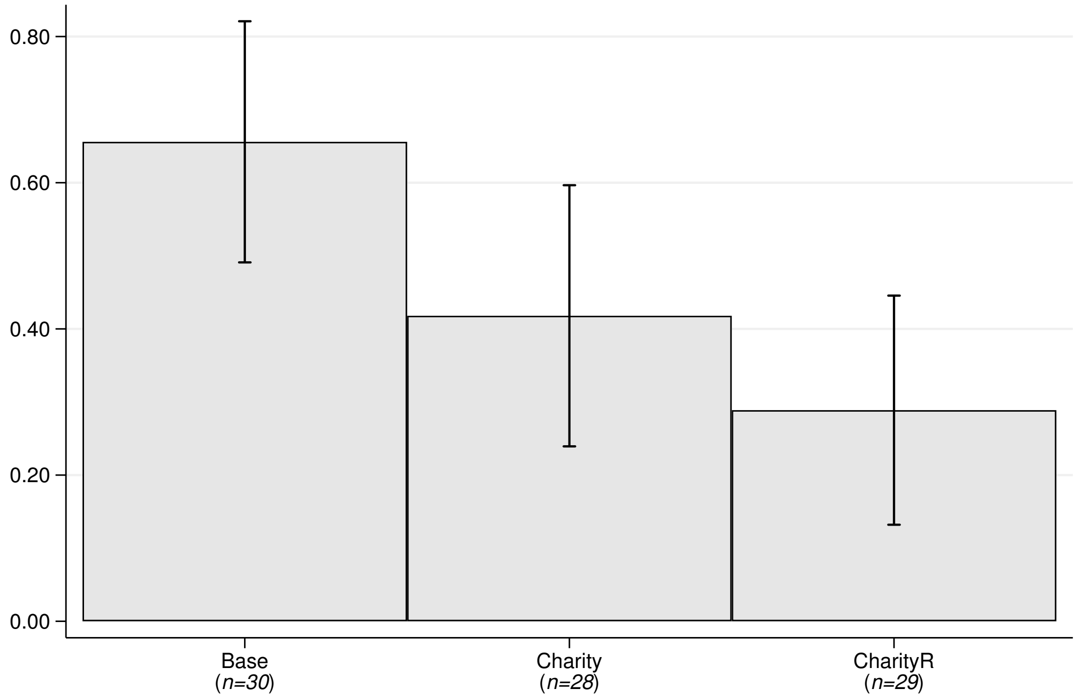


Figure 5: Mean and 95% confidence interval of the proportion of messages for over-reporting in Part II of the group treatments.

treatments, and both research assistants assigned the same number in all coded messages. A chat message then becomes a message for over-reporting when the coded number is higher than the observed die roll. For each group, we finally compute the proportion of messages for over-reporting.

In the chat, group members often agree on whether to over-report: the proportion of messages for over-reporting is 0 or 1 in 77%, 78% and 72% of groups in the *G-Base*, *G-Charity* and *G-CharityR* treatments. As can be expected, there is a positive and significant correlation between the proportion of messages for over-reporting and groups over-reporting in Part II in the *Base* (Spearman's $\rho = 0.883$, $p < 0.001$), *Charity* ($\rho = 0.921$, $p < 0.001$) and *CharityR* ($\rho = 0.852$, $p < 0.001$) treatments.

Figure 5 details the mean and 95% confidence interval of the proportion of messages for over-reporting for groups in the different treatments. We observe that this proportion is higher in the *Base* treatment relative to the *Charity* ($U = 313.5$, $p = 0.078$) and *CharityR* ($U = 254.5$, $p = 0.004$) treatments. In Table 2, we use the fractional logit model (e.g., Papke and Wooldridge, 1996, 2008) to further control for the die number observed in Part II as well as the proportion of group members who over-reported in Part I. The regressions show that the proportion of messages

Table 2: Fractional Logit regression estimates: How the proportion of messages for over-reporting differs across groups in the group treatments.

Dependent Variable: Proportion of messages for over-reporting		
Regression	(1)	(2)
Reference group: Groups in the *G-Base* treatment.		
G-Charity	-1.38** (0.58)	-1.71** (0.68)
G-CharityR	-2.07*** (0.58)	-2.11*** (0.69)
Points observed (Part II)		-0.64*** (0.17)
♯ of members over-report (Part I)		0.53* (0.29)
Constant	1.51*** (0.44)	2.92*** (0.89)
n	69	69
Pseudo R^2	0.12	0.22
$\chi^2(1)$: *G-Charity* = *G-CharityR*	1.59	0.59

Notes. Instances where groups observed 5 points omitted.
***, ** and * denote $p < 0.01$, $p < 0.05$ and $p < 0.10$.

for over-reporting is significantly higher for groups in the *Base* treatment than in the *Charity* ($p \leqslant 0.018$) and the *CharityR* ($p < 0.001$) treatments. There are no significant differences between groups in the *Charity* and the *CharityR* treatments ($p \geqslant 0.208$). This leads us to the next result:

Result 4. *In the chat, groups exchange fewer messages that call for over-reporting when they know that over-reporting harms the charity.*

These findings suggest that groups over-report more than individuals in the *Base* treatments but not the *Charity* and *CharityR* treatments because group members do not voice arguments for dishonesty when the charity is involved.[6]

3.3 Spillovers of dishonesty

Finally, we study behaviour across parts to see whether dishonesty can spillover from one part to the next.

Do dishonest individuals make groups more dishonest? We first study whether over-reporting individuals make groups over-report more. We do not find any

[6]We also looked at the deliberations from the individual treatments. There is, however, very little to exploit, as most subjects simply restated the instructions; few deliberations can be meaningfully categorised. We give more information in Appendix D.

significant between-treatment differences in the number of group members who over-reported in Part I for groups in the *G-Base*, *G-Charity* and *G-CharityR* treatments (Kruskal-Wallis test, $p = 0.202$), so group composition does not seem to explain the above results.

Table 2 already showed us that there is a positive and significant relationship between the number of group members who over-reported in Part I and the proportion of chat messages that call for over-reporting.

This relation, however, does not seem to translate to actions: we find no significant correlation between the number of over-reporting subjects in Part I and over-reporting in Part II for groups in the *G-Base* (Spearman's $\rho = 0.235$, $p = 0.210$), *G-Charity* ($\rho = 0.117$, $p = 0.553$) and *G-CharityR* ($\rho = -0.117$, $p = 0.544$) treatments. The regression in Appendix B.1.2 confirms this result while further controlling for points observed in Part II. Note that this holds even when we look at groups for which all members over-reported or did not over-report at all in Part I.

We therefore have the following result:

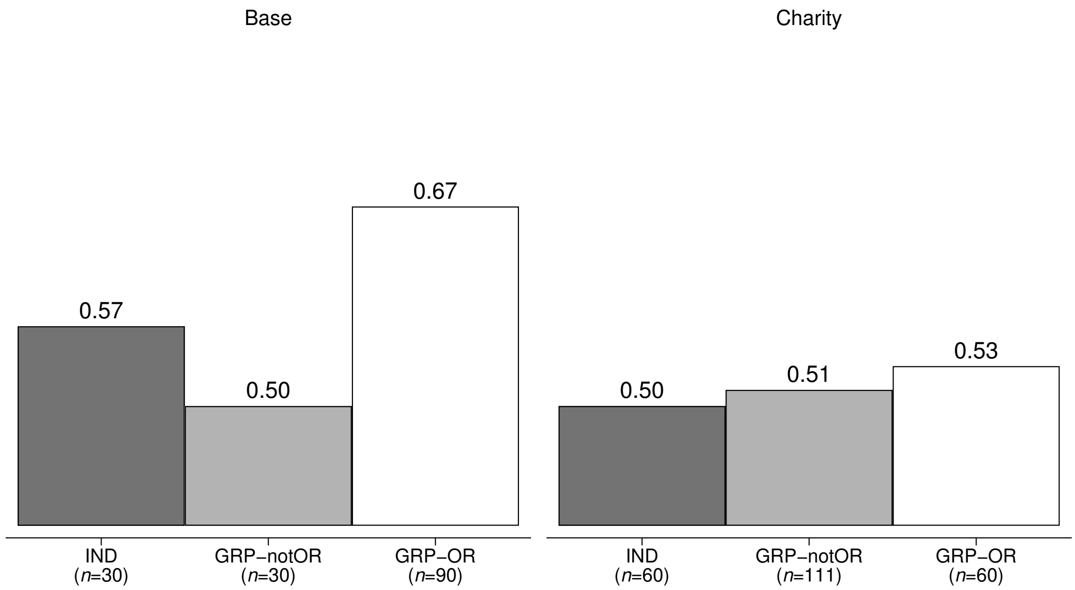
Result 5. *The number of over-reporting individuals within the group has only a limited influence on the group's decision to over-report. The same is true even when the charity is involved.*

Do dishonest groups make individuals more dishonest? Finally, we look at whether subjects who were in a group that over-reported in Part II over-report themselves in Part III. For this we compare subjects who were in a group that over-reported in Part II (GRP-OR subjects) to subjects who were in a group that did not over-report in Part II (GRP-notOR subjects) and to subjects who were not in a group in Part II (IND subjects).

Figure 6 shows the proportion of IND, GRP-notOR and GRP-OR subjects who over-report in Part III. We pool the *Charity* and *CharityR* treatments together since subjects in these treatments faced the same situation in Part III. We see that the proportion of over-reporting GRP-OR subjects is only marginally higher than of IND and GRP-notOR subjects.

The regressions in Table 3 look at the impact of being in an over-reporting or a non-over-reporting group in Part II on the likelihood of over-reporting in Part III, controlling for behaviour in Part I, gender and points observed in Part III.[7]

[7] A reviewer pointed out that, since GRP-OR in Part II and Over-report in Part I might be related, models (2) and (4) might be suffering from multicollinearity. Result 5 showed us that the number of subjects who over-reported in Part I has very limited influence on whether a group over-reports in Part II, so even if there is collinearity we know it must be limited. To be sure, we use the VIF (Variance Inflation Factor) and obtain main values of 1.07 (*Base*), 1.02 (*Charity*) and 1.04 (*CharityR*),



Notes. All subjects made individual decisions in Part III. IND: subjects who made individual decisions in Part II. GRP-notOR (GRP-OR): subjects who made group decisions in Part II and whose group had not (had) over-reported.

Figure 6: Proportion of IND, GRP-notOR and GRP-OR individuals who over-report in Part III.

The estimates show that GRP-OR subjects are significantly ($p \leqslant 0.053$) more likely to over-report in Part III relative to IND and GRP-notOR subjects in the *Base* and *Charity* treatments. Also, we do not find the likelihood of over-reporting to be significantly higher for GRP-notOR relative to IND subjects. This leads us to the following result:[8]

Result 6. *Individuals are more likely to over-report if they were previously in a group that over-reported, irrespective of whether over-reporting hurt the experiment or the charity.*

4 Dishonesty or pro-sociality?

We have seen that, with the charity, groups are not more likely to over-report than individuals. In fact, when they do, they over-report to a lesser extent. As Kocher et al. (2018) did, we interpret over-reporting as dishonesty and so conclude that

which suggests there is little multicollinearity, if any.

[8]We find the same result using the data from Kocher et al. (2018): subjects from over-reporting groups are significantly more likely to over-report than subjects who were not in a group ($p = 0.021$) and subjects from non-over-reporting groups ($p < 0.001$). Subjects from non-over-reporting groups report similarly as subjects who were not in a group ($p = 0.124$).

Table 3: Logit regression estimates: The spillovers from membership of over-reporting groups.

Dependent Variable: Over-report in Part III.	BASE		CHARITY	
Regression	(1)	(2)	(3)	(4)
Reference group: IND subjects				
GRP-notOR (Part II)	-0.66	-0.81	0.21	-0.45
	(0.58)	(0.63)	(0.38)	(0.49)
GRP-OR (Part II)	1.19*	1.45**	1.01**	1.64***
	(0.62)	(0.66)	(0.43)	(0.56)
Points observed (Part III)		-0.41**		-0.75***
		(0.18)		(0.12)
Over-report (Part I)		-2.04**		2.97***
		(0.88)		(0.52)
Constant	0.88*	1.31*	0.51*	0.82*
	(0.45)	(0.72)	(0.29)	(0.47)
n	96	96	186	186
Pseudo R^2	0.09	0.20	0.03	0.33
$\chi^2(1)$: GRP-OR = GRP-notOR	11.11***	12.42***	9.50***	9.05***

Notes. Instances where subjects observed 5 points are omitted.
Standard errors clustered at the group level.
***, ** and * denote $p < 0.01$, $p < 0.05$ and $p < 0.10$.

groups are as dishonest as individuals when the charity is involved.

A further analysis of the chat data corroborates this interpretation. After the *coding* of the chat messages, where they assigned a recommended number to each chat message, the research assistants also *categorised* chat messages based on the types of arguments subjects used in the chat to justify their proposal (again, see Appendix C for details). Subjects rarely revealed the motivation behind their proposed numbers ("we should report the number we saw" or "we should give a lot to the charity"), so only about 20% of the chat messages could be categorised.[9] With this caveat in mind we find that the proportion of arguments in favour of honesty goes from 33% in the *G-Base* treatment to 56% in the *G-Charity* treatment and 43% in the *G-CharityR* treatment.

At the same time, however, the proportion of statements in favour of pro-sociality jumps from 0% in the *G-Base* treatment to 36% in the *G-Charity* treatment and 51% in the *G-CharityR* treatment. It is thus possible that groups do not become less dishonest with the charity, but become more pro-social. They would over-report less to give more to the charity, but not because of dishonesty. As a

[9]Around 100%, 94%, and 97% of categorised chats in the *Base*, *Charity*, and *CharityR* treatments are also coded.

result, innate dishonesty of groups and heightened pro-sociality in the presence of the charity would cancel each other out, leading to no difference in over-reporting between groups and individuals. To test for this alternative explanation, we designed a series of follow-up treatments: the *Dictator* treatments.

4.1 Procedure

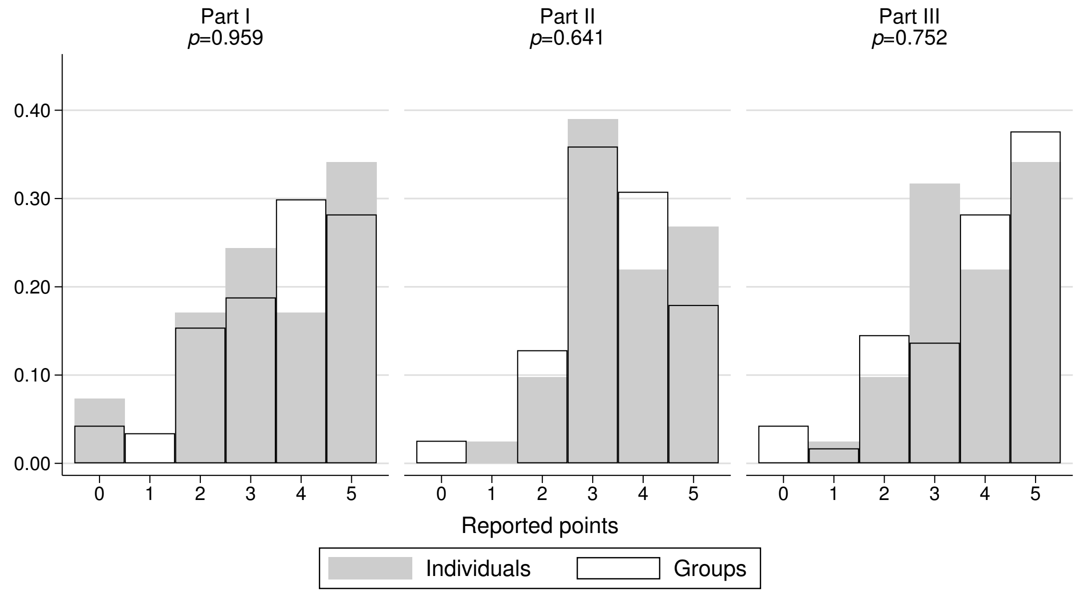
The *Dictator* treatments keep the same three-part structure of the *Charity* treatments but remove the die roll: subjects report a number that determines points for themselves and for the charity, still as described by Table 1. Therefore, in Part I and III all subjects play the standard dictator game with the charity as recipient. In Part II, subjects in the *I-Dictator* treatment continue to play the standard dictator game while those in the *G-Dictator* treatment play the dictator game as a group: all group members independently report an amount for the group and the allocation is implemented only if they all report the same amount. By focusing on Part II and comparing individuals in *I-Dictator* and groups in *G-Dictator*, we can test whether groups are indeed more pro-social than individuals.

The sessions took place online, using zTree Unleashed (Duch et al., 2020), in May and June 2021. The move to online sessions meant we had to pay subjects via bank transfer several weeks later. This change, however, affects both the *I-Dictator* and the *G-Dictator* treatment, and so should not affect the results when comparing these two treatments. Given the number of changes—no die roll, online, different payment—we will not compare directly the new *Dictator* treatments and the old *Base*, *Charity*, and *CharityR* treatments.

The instructions can be found in Appendix E.3. 43 subjects participated in the *I-Dictator* treatment and 126 in the *G-Dictator* treatment. After removing subjects who dropped out during the experiment—and, in the group treatment, groups in which at least one group member dropped out—we are left with 41 subjects in *I-Dictator* treatment and 117 subjects (39 independent observations) in the *G-Dictator* treatment.

4.2 Results

Figure 7 shows the distribution of points reported in each Part by individuals and by groups. The reported points are the amount that individuals or groups intend to keep for themselves. In Parts I and III, subjects make individual decisions and so are identical in the *I-Dictator* and the *G-Dictator* treatments; we find no significant between-treatment differences for these parts ($U > 2386$, $p \geqslant 0.754$). In



Note. p-values from Mann-Whitney U tests comparing the reported points of individuals and groups.

Figure 7: Histogram of points reported by individuals and groups in Part II of the *Dictator* treatment.

Part II, individuals report slightly higher points than groups, but the difference is not significant (U = 751.5, p = 0.639). Therefore, individuals and groups behaved the same in the *Dictator* treatments.

We also follow the methodology from Cason and Mui (1997) and classify individuals and groups as *selfish* or *pro-social* types. We choose the mean number of points reported in Part I, 3.5 points, as the neutral number of points.[10] Individuals or groups that reported points above 3.5 are categorised as selfish types, and those below, as pro-social types. Around 51% of individuals and groups are classified as pro-social types in Part II ($\chi^2 = 1.2927$, p = 0.524). We also see that selfish individuals report slightly higher points than selfish groups, but the difference is not significant (U = 155.5, p = 0.262). Taken together these observations lead us to a new result:

Result 7. *In the Dictator treatments, groups are as pro-social as individuals.*

This result indicates that what we observed in the previous treatments—groups over-reporting more than individuals in the *Base* treatments, but not doing so when the charity is involved—cannot be explained by groups being more pro-social

[10]The conclusions do not change if we use the median (4 points).

in the presence of the charity. Instead, Result 7 points to groups behaving less dishonestly with the charity as being the prime explanation of our findings.

In Appendix B.2 we also look at how parts affect each other in the *Dictator* treatment. We find that, while selfish members make groups more selfish, being in a selfish or in a pro-social group has only a limited impact on subsequent reports.

Finally, we cannot exploit the chat data in the *Dictator* treatments: groups communicate very little, agreement on a number to report is quick, and subjects almost never give reasons behind their actions. As a result, the research assistants could only categorise 10 chat messages.

5 Post-study probability estimates

Before concluding, to assess the robustness of our findings we follow Maniadis et al. (2014) and compute the post-study probability (PSP): 'the probability that a declaration of a research finding, made upon reaching statistical significance, is true.' Ideally, we want PSP > 0.5. The PSP depends on our prior π, on power $1 - \beta$, on the significance level α, and on the number k of independent research teams who work on the same problem.

In Figure 8 we fix $\alpha = 0.05$ and plot the PSP for different values of these parameters. We observe that power has only a limited impact on the PSP. This observation is reassuring given our small sample size.

The prior measures the probability before running the experiment that the effect we were looking for—groups not being more dishonest than individuals—is true. It is unclear which prior would be reasonable to assume in our case, but note that the PSP is always greater than 0.5 as long as the prior is greater than 0.20, no matter the power and the number of competing research teams. These numbers suggest that we can expect our results to be replicated.

6 Conclusion

We report the results of a laboratory experiment testing whether groups are more dishonest than individuals. We replicate the study by Kocher et al. (2018) but also show that, when dishonesty harms a local charity and not the experimenter, groups are not more dishonest than individuals. In fact, groups can even help moderate the extent of the dishonesty.

We gave information about the charity only in the *Charity* and *CharityR* treatments. Mentioning the charity might prime moral actions, and since we did

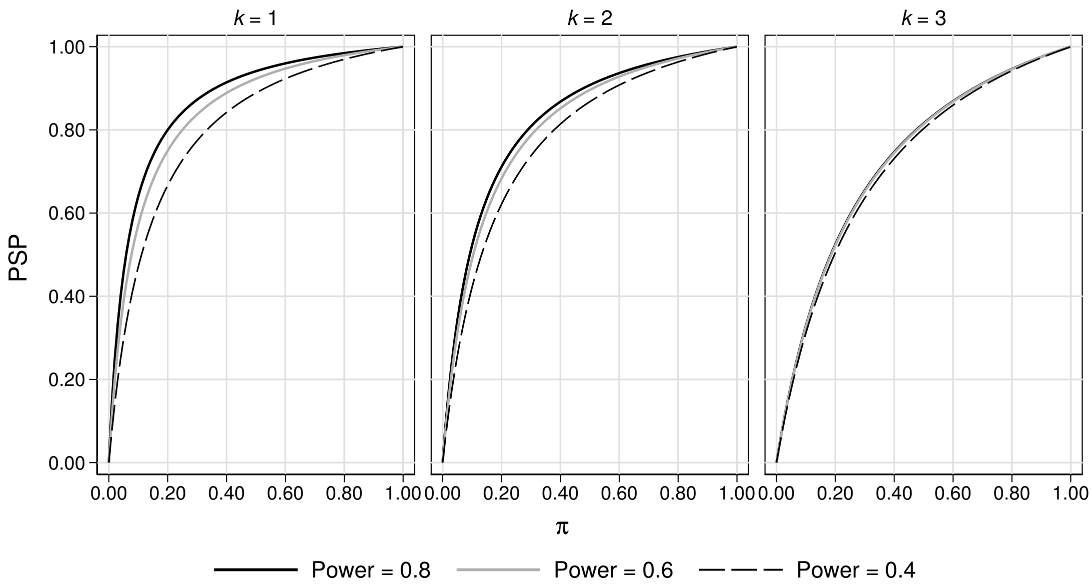


Figure 8: Post-study probability computed for different levels of the prior π, power $1 - \beta$, and the number k of competing research teams (assuming $\alpha = 0.05$).

not mention the charity in the *Base* treatments, this difference might explain our results. Note, however, that we do not observe any difference for individuals in Parts I and II between the *Base*, *Charity*, and *CharityR* treatments. Therefore, if there is priming, it must be limited, or it must manifest itself only in the group setting to explain our findings. We feared that mentioning the charity in the *Base* treatment, where it plays no role whatsoever, could generate confusion. It would also have been hard to present the charity without creating experimenter demand effects and making the subjects think they should use this information.

We have addressed *whether* groups are more dishonest than individuals with a charity—they are not—but we have only partially found *why*. Following Kocher et al. (2018) and the analysis of the chat data, we favour a social image concerns explanation. With the *Dictator* treatments, we have ruled out an explanation based on social preferences. But there are other explanations; subjects in treatments involving the charity could for example form beliefs about the social preferences of others. We did not design our experiment to precisely tease out these explanations, and we leave this work to future research. To work on the social image channel, one could make groups come on stage and read a copy of their transcript. To work on the beliefs channel, one could directly measure beliefs, or control for them by providing subjects with the decisions the other group members did in Part I.

Our results imply that organisations should make explicit who is getting hurt if employees engage in dishonest behaviour. Employees could be tempted to coordinate to extract money from their company, for example by submitting and countersigning fake bills. The company may preemptively signal that such practices hurt the company's profits and ultimately the employees themselves. If the company is public, it could even signal that it is funded from taxes that the employees themselves pay.

The diffusion of responsibility literature (Darley and Latàn, 1968) has instead shown that people are less likely to stop dishonest behaviour if they are in a group, which results in groups being more dishonest than individuals. Note, however, that in our experiments the default is being honest—reporting the die number seen—and being dishonest requires groups to coordinate on deviating from this honest default. On the other hand, with diffusion of responsibility the default is often dishonesty. For example, in Choo et al. (2019) groups have to coordinate to whistleblow and report dishonest behaviour; the dishonest outcome happens if they do not act. They find that indeed groups are less likely to whistleblow than individuals. Studying the impact of changes in the status-quo on dishonesty is something we leave to future research.

Our results also highlight that dishonesty is context-sensitive. A potential follow-up to our experiment would be to keep the experimenter as a victim of subjects' dishonesty but make it explicit, for example by stressing who funds the study. Future research could also look for other cues that similarly affect dishonesty.

References

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond, "Preferences for Truth-Telling," *Econometrica*, 2019, *87* (4), 1115–1153.

Barr, Abigail and Georgia Michailidou, "Complicity without Connection or Communication," *Journal of Economic Behavior & Organization*, October 2017, *142*, 1–10.

Cason, Timothy N. and Vai-Lam Mui, "A Laboratory Study of Group Polarisation in the Team Dictator Game," *The Economic Journal*, September 1997, *107* (444), 1465–1483.

Choo, Lawrence, Veronika Grimm, Gergely Horváth, and Kohei Nitta, "Whistleblowing and Diffusion of Responsibility: An Experiment," *European Economic Review*, 2019, *119*, 287–301.

Chytilová, Julie and Václav Korbel, "Individual and Group Cheating Behavior: A Field Experiment with Adolescents," Technical Report, IES Working Paper 2014.

Conrads, Julian, Bernd Irlenbusch, Rainer Michael Rilke, and Gari Walkowitz, "Lying and Team Incentives," *Journal of Economic Psychology*, 2013, *34*, 1–7.

Darley, John M. and Bibb Latàn, "Bystander Intervention in Emergencies: Diffusion of Responsibility," *Journal of Personality and Social Psychology*, 1968, *8* (4), 377–383.

Drouvelis, Michalis and Graeme Pearce, "Understanding the Link between Intelligence and Lying," SSRN Scholarly Paper ID 3898321, Social Science Research Network, Rochester, NY 2021.

Duch, Matthias L., Max R. P. Grossmann, and Thomas Lauer, "Z-Tree Unleashed: A Novel Client-Integrating Architecture for Conducting z-Tree Experiments over the Internet," *Journal of Behavioral and Experimental Finance*, December 2020, *28*, 100400.

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

_ **and Franziska Föllmi-Heusi**, "Lies in Disguise – An Experimental Study on Cheating," *Journal of the European Economic Association*, 2013, *11* (3), 525–547.

Gino, Francesca, Shahar Ayal, and Dan Ariely, "Self-Serving Altruism? The Lure of Unethical Actions That Benefit Others," *Journal of Economic Behavior & Organization*, 2013, *93*, 285–292.

Gneezy, Uri, "Deception: The Role of Consequences," *American Economic Review*, March 2005, *95* (1), 384–394.

Greiner, Ben, "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE," *Journal of the Economic Science Association*, 2015, *1* (1), 114–125.

Kocher, Martin G., Simeon Schudy, and Lisa Spantig, "I Lie? We Lie! Why? Experimental Evidence on a Dishonesty Shift in Groups," *Management Science*, September 2018, *64* (9), 3995–4008.

Maggian, Valeria, "Chapter 3.3 - Negative Externalities of Cheating: An Experiment with Charities," in Alessandro Bucciol and Natalia Montinari, eds., *Dishonesty in Behavioral Economics*, Perspectives in Behavioral Economics and the Economics of Behavior, Academic Press, January 2019, pp. 183–191.

Maniadis, Zacharias, Fabio Tufano, and John A. List, "One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects," *American Economic Review*, January 2014, *104* (1), 277–290.

Muehlheusser, Gerd, Andreas Roider, and Niklas Wallmeier, "Gender Differences in Honesty: Groups versus Individuals," *Economics Letters*, 2015, *128*, 25–29.

Papke, Leslie E. and Jeffrey M. Wooldridge, "Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates," *Journal of Applied Econometrics*, 1996, *11* (6), 619–632.

_ **and** _ , "Panel Data Methods for Fractional Response Variables with an Application to Test Pass Rates," *Journal of Econometrics*, 2008, *145*, 121–133.

Pruckner, Gerald J. and Rupert Sausgruber, "Honesty on the Streets: A Field Study on Newspaper Purchasing," *Journal of the European Economic Association*, 2013, *11* (3), 661–679.

Schram, Arthur and Gary Charness, "Inducing Social Norms in Laboratory Allocation Choices," *Management Science*, 2015, *61* (7), 1531–1546.

Soraperra, Ivan, Ori Weisel, Ro'i Zultan, Sys Kochavi, Margarita Leib, Hadar Shalev, and Shaul Shalvi, "The Bad Consequences of Teamwork," *Economics Letters*, 2017, *160*, 12–15.

Sutter, Matthias, "Deception Through Telling the Truth?! Experimental Evidence from Individuals and Teams," *The Economic Journal*, January 2009, *119* (534), 47–60.

Weisel, Ori and Shaul Shalvi, "The Collaborative Roots of Corruption," *Proceedings of the National Academy of Sciences*, 2015, *112* (34), 10651–10656.