

# On the interpretation of bribery in a laboratory corruption game: moral frames and social norms

Ritwik Banerjee<sup>1</sup>

Received: 25 August 2013 / Revised: 27 February 2015 / Accepted: 9 April 2015 /

Published online: 5 May 2015

© Economic Science Association 2015

**Abstract** Past studies on laboratory corruption games have not been able to find consistent evidence that subjects make “immoral” decisions. A possible reason, and also a critique of laboratory corruption games, is that the experiment may fail to trigger the intended immorality frame in the minds of the participants, leading many to question the very *raison d’être* of laboratory corruption games. To test this idea, we compare behavior in a harassment bribery game with a strategically identical but neutrally framed ultimatum game. The results show that fewer people, both as briber and bribee, engage in corruption in the bribery frame than in the alternative and the average bribe amount is lesser in the former than in the latter. These suggest that moral costs are indeed at work. A third treatment, which relabels the bribery game in neutral language, indicates that the observed treatment effect arises not from the neutral language of the ultimatum game but from a change in the sense of entitlement between the bribery and ultimatum game frames. To provide further support that the bribery game does measure moral costs, we elicit the shared perceptions of appropriateness of the actions or social norm, under the two frames. We show that the social norm governing the bribery game frame and ultimatum game frame are indeed different and that the perceived sense of social appropriateness plays a crucial role in determining the actual behavior in the two frames. Furthermore, merely relabelling the bribery game in neutral language makes no difference to the social appropriateness norm governing it. This indicates that, just as in the case of actual behavior, the observed difference in social appropriateness norm between

---

**Electronic supplementary material** The online version of this article (doi:[10.1007/s10683-015-9436-1](https://doi.org/10.1007/s10683-015-9436-1)) contains supplementary material, which is available to authorized users.

---

✉ Ritwik Banerjee  
rbanerjee@econ.au.dk

<sup>1</sup> Aarhus University, Aarhus, Denmark

bribery game and ultimatum game comes from the difference in entitlement too. Finally, we comment on the external validity of behavior in lab corruption games.

**Keywords** Corruption · Framing effects · Social norms · External validity

**JEL Classification** C91 · C92 · D03

## 1 Introduction

Laboratory experiments in Economics, despite their widespread use today, have often been subjected to (unfair) criticism on account of lacking realism and generalizability. In particular, critics have questioned the external validity of the results and have argued that inconsequential size of stakes, desire to maintain a positive self image and a lack of an appropriate moral frame, among other things, confound the inferences drawn from them [for a general critique of lab experiments see Levitt and List (2007)]. However, lab experiments have since received considerable defense<sup>1</sup> as economists have concurred that the possibility of controlled variations in lab environments, which makes causal inferences possible, is a valuable source of knowledge (Falk and Heckman 2009).

This is reflected in the increasing popularity of lab based corruption games as tools to identify, measure and understand the otherwise clandestine phenomenon called corruption. These tools allow us to manipulate the institutional (Alatas et al. 2009; Abbink et al. 2014; Banuri and Eckel 2012b), social (Banerjee et al. 2015) and cultural contexts (Cameron et al. 2009; Barr and Serra 2010) which in turn help us draw causal inferences. However, it is not entirely surprising that laboratory corruption games have been received with a fair share of skepticism too, primarily on the following two grounds. First, external validity of the results may be affected since subjects may maintain a positive self image in lab and second, experimenters are unable to impose an appropriate moral frame in an artificial lab environment. While it is possible that the desire to maintain self image may be a confound, any observed measure of corruption in the lab should be interpreted as a lower bound of behavior in the field. Furthermore, if it is the treatment effect, as opposed to the absolute level, that is of interest, then a laboratory corruption game is a useful and relatively inexpensive tool for the researchers.<sup>2</sup> Finally, Armantier and Boly (2013) in a clever and convincing study find evidence that results from lab based corruption games are indeed generalizable outside the lab and that the generalizability holds across cultures.

---

<sup>1</sup> For instance, Camerer (Forthcoming) in his reply to Levitt and List (2007) shows that the overwhelming majority of lab experiments do indeed generalize to comparable field settings. Kessler and Vesterlund (2014) emphasize that it is the qualitative effect (i.e. direction of effect), as opposed to the quantitative effect (i.e. precise size of the effect), which is more generalizable.

<sup>2</sup> Levitt and List (2007) warn us that lab experiments may not reveal so much when it comes to identifying deep structural parameters. Then one needs to take recourse to field experiments despite the fact that it is often prohibitively expensive as List (2012) found out with a \$103,000 field experiment on corruption.

However, the other important feature of lab based corruption games is yet to be established on firm grounds. Do subjects in the game actually make decisions under the moral frame that the experimenter intends to impose? In other words, for an act to be interpreted as an unethical act we must find compelling evidence of a moral cost at play in the exercise of that act. This important feature, interpreted in terms of the treatment effect between the corruption frame and a neutral frame, has either not been found in the past studies (Abbink and Hennig-Schmidt 2006) or has been found in some cases while not in others (Barr and Serra 2009).<sup>3</sup>

While all studies incorporating such a game show that subjects do not opt for the highest material pay off, it is still not clear whether they act under the influence of the psychological frame of morality that the experimenter intends to impose. In this paper we revisit the question whether moral costs are at play in a lab based corruption game. We implement a design where in one treatment a harassment bribery game is implemented. In the counterfactual, a strategically identical but differently framed ultimatum game is implemented such that the psychological reference point of the subjects is changed by a subtle change in entitlement. Since the treatment and the counterfactual differ in terms of framing and entitlement, we implement a third treatment in order to identify whether it is the framing or the entitlement which leads to the observed treatment effect.

As a second test of whether the appropriate moral frame is at work in these corruption games, we elicit measures of how socially appropriate actions in each of the three treatments are. To do so we follow Krupka and Weber (2012) and elicit commonly held beliefs about social norms<sup>4</sup> through a coordination game designed to elicit second order beliefs in an incentive compatible way. Socially held beliefs about what is the right thing to do potentially play a vital role in determining actual behavior. The author recalls an instance where a friend from India underwent a radical change in his driving practices after a four week travel to the United States, as he followed the newly internalized norms (e.g. sticking to lanes, not honking). His action was being dictated, at least in part, by what was considered appropriate in the US. The strategy of eliciting measures of social appropriateness therefore enables us to see whether participants assign different moral standards for the two strategically identical but differently framed environments. This, in turn, helps us explore in very concrete terms, how the elicited measures of social appropriateness affect actual behavior, if at all. In our between subject design “citizens” and “public

<sup>3</sup> Two distinct questions emerge from this: One, do people operate under different moral environment in the two frames and if not, is this why one does not observe framing effects? Two, what is the appropriate counterfactual of a corruption experiment that may potentially yield evidence of moral cost at work? Abbink and Hennig-Schmidt (2006) speculate that neutral frames are insufficient to induce an alternative behavioral norm.

<sup>4</sup> We follow Fehr and Fischbacher (2004) to define social norm for our purpose: “The standards of behavior that are based on widely shared beliefs how individual group members ought to behave in a given situation”. There have been various other definition of social norms in the literature: Pareto noted that “...people have opinions about how they should or should not behave. They also have opinions about how others should or should not behave.” Ostrom (2000) further emphasized on the mutually shared aspect of social norms and defined it as “...shared understandings about actions that are obligatory, permitted or forbidden”.

officials” play a real effort harassment bribery game.<sup>5</sup> A citizen performs a task and earns a prize if successful. A public official, however, may want a bribe in order to let the citizen have her prize. The citizen may then accept or reject the demand for a bribe. In the counterfactual treatment, Participant A (analogous to the Citizen)—upon successfully completing the task—earns the right to go to the second stage of the game. At the second stage, Participant B (analogous to the Public Official) plays an ultimatum game, with the same stake size as the prize of the bribery frame, and decides how much to share with Participant A, if anything. The latter can then accept or reject Participant B’s offer. Though strategically identical, these two treatments differ in terms of the language and the sense of entitlement among the subjects. To understand whether the observed difference comes from the difference in the entitlement or the language, we conduct a third treatment which retains the bribery game’s sense of entitlement but differs only in terms language (i.e. it uses a neutral language). Finally, we relate corruption behavior in the framed lab experiment to stealing behavior, in order to understand what the behavioral primitive measured by corruption in bribery game actually is. In a separate experiment, we elicit the social appropriateness of the frames and then examine whether choice of action in one frame is considered as more socially inappropriate than strategically identical choice of action in the other.

Our results show that retained shares in the ultimatum game are higher than the bribes demanded under the corruption game. However, there is no difference in the bribe amount demanded in the bribery game played in loaded language and that played in neutral language. Thus merely changing the language from loaded to neutral makes little difference in behavior in our set up but changing the psychological reference point by altering the sense of entitlement does. These results point to a psychological moral cost at work in the bribery frame—evidence of which has been mixed in the past studies. However, there is a caveat—unlike in past studies we implement a two-player game where corruption takes the form of a harassment bribe and does not generate negative externalities to others. Besides, our experiment is conducted in India which is highly corrupt and not in western countries with low corruption norms.

In addition to difference in actual behavior, social appropriateness measure differs considerably for the same strategies across the bribery game and ultimatum game treatments. Strategies are considered more socially inappropriate in the corruption game than in the alternative. Once again we find no difference in social appropriateness norms between the bribery game in loaded and neutral language, indicating that a change in language by itself does not create an alternative moral environment but change in entitlement does. Finally, we show that the observed actions can be explained by the elicited social appropriateness measures. Overall, the findings suggest that in order for laboratory corruption games to measure unethical behavior, they should have an entitlement component and not just use

<sup>5</sup> Harassment bribery is a form of bribery where a public official asks for bribe from a citizen who is entitled to a service that the official is obligated to provide. This form of bribery is very common in developing countries where citizens are entitled to government services but either they have to pay a bribe in order to obtain them or avoid inordinate procedural delays. Such services include issuance of a passport or a driver’s license (given that the candidate has passed the driving test).

corruption loaded language. If they do so then they lend themselves as cheap diagnostic tools which can help analyze various public policy interventions.

The contributions of this paper are the following. First, the paper makes a contribution to the framing literature by studying the effects of differences in entitlement and language in three strategically identical frames. Second, the treatment effects we find indicate that the experimenters can successfully impose the appropriate moral frame in laboratory corruption games if subjects, who are being demanded bribe of, feel entitled to their earning. This is indeed good news for laboratory experiments in general and lab based corruption games in particular. Our results, combined with the results obtained by Armantier and Boly (2013), will go a long way in convincing the skeptics of lab based corruption games. However, our results also indicate that a loaded language frame may be neither necessary nor sufficient to create an immoral environment in a bribery game. What is more important is inducing a sense of entitlement among those players from whom bribe is being extracted. This may have been one reason why past studies did not uniformly find evidence of framing effect—the alternate psychological reference point may not have been created in the counterfactuals. Third, we measure in clear categorical terms the amorphous concept of norms governing an unethical act and contrast it with that of a strategically identical but differently framed game. Fourth, to the best of our knowledge, this is one of the first attempts to answer how actual corruption behavior is related to socially held perceptions about what is morally the right thing to do. Notice that this result is distinct from other prevalent explanations of why people obey norms—e.g. fear of sanctions upon violation of norms (Fehr and Fischbacher 2004), to obtain self esteem (Bernheim 1994) or for informational advantages (Banerjee 1992).

This paper is divided into six sections. Following the introduction, Sect. 2 briefly reviews the literature of framing effects as well as that of social norms. In Sect. 3 the experimental designs of the main treatments are laid out along with the design for elicitation of social appropriateness. Section 4 presents and discusses the main results while the summary and conclusion is given in Sect. 5.

## 2 Related literature

The literature, on how perceptions about what is morally the right thing to do affect actual corrupt behavior, is still nascent. Not only are such perceptions vulnerable to common factors which might also affect corrupt behavior, but conversely corrupt behavior might determine such perceptions as well. Nonetheless, economists have turned to cross country studies to relate cultural norms to corruption<sup>6</sup> [see for instance Treisman (2000), Serra (2006) and Fisman and Miguel (2007)], though lack of comparable data greatly limits that scope.

<sup>6</sup> Though here we mainly discuss the experimental literature, several theoretical explanations have been advanced about why people may conform to social norms. While Banerjee (1992) explained preponderance of adherence to social norms in terms of informational advantage of others, Fehr and Fischbacher (2004) and Bernheim (1994) explained the conformity to social norms in terms of fear of sanction in case of violation and desire to gain social esteem, respectively.

This limitation has been tackled in the recent years by introducing controlled variations in the experiments but again with mixed success [for a review of laboratory corruption games see Serra and Wantchekon (2012) and for corruption and culture see Banuri and Eckel (2012a)]. Cameron et al. (2009) conduct comparable experiments in countries with different corruption levels and Barr and Serra (2010) study corruption behavior among students from different countries but both fail to find across the board effect of home country corruption indices on behavioral responses in the lab. However, social norms in these studies are proxied by the aggregate indices of corruption, which measures (however inaccurately) perceptions of prevalence rather than that of appropriateness.

From a different perspective, behavior has been analyzed through the lens of contagion effect or the tendency to conform to behavior of others (see for instance Bicchieri and Xiao (2009), Krupka and Weber (2009) and Cason and Mui (1998) for dictator games and Innes and Mitra (2013) for deception games). Conceptually the prevalence of unethical acts is very different from social norms which are a set of commonly held, unwritten normative codes governing our actions. Despite these codes not binding, they impose a substantial cost on people by way of inflicting moral sanctions for following a contrarian way.

Accordingly in our view social norms, culture, prevalence though often approximated with each other, are very different. This approximation is tackled in this paper by directly eliciting measures of social norm or appropriateness using a coordination device introduced by Krupka and Weber (2012). We use this device to elicit the measure of social appropriateness for the alternative frames we study and then predict cross treatment behavior.<sup>7</sup> In the context of laboratory corruption games the role of framing has been analyzed by Abbink and Hennig-Schmidt (2006) and then subsequently by Barr and Serra (2009). Abbink and Hennig-Schmidt (2006) rely on using value free or “neutral” vocabulary to create a context free environment and pit it against a “loaded” environment. Their hypothesis is that the moral frame in a “loaded” environment will lead to less corruption. However, they do not find a framing effect and attribute it, among other things, to the fact that even with a neutral frame, the game captures the quintessential features of corruption.

Barr and Serra (2009) later point to several other plausible reasons why no framing effect was found—mismatch between the created briber-bribee environment and real life experiences of subjects, triggering of role play rather than moral compunctions induced by “artificial” environments, punishment strategies in both neutral and loaded versions, etc. However, despite correcting for the majority of these limitations in their own study, Barr and Serra (2009) find mixed results as far as framing effects are concerned. For low externalities, private citizens are not found to be more likely to offer bribes under a corruption frame but a statistically significant difference is found in the mean bribe offered. However, for high externalities, and also for the pooled sample, private citizens are more likely not to offer a bribe in the corruption frame than in the abstract frame; though no significant

---

<sup>7</sup> Though Reuben and Riedl (2013) acknowledge that this tool is clearly a superior method of eliciting normative views than those based on questionnaires, only a handful of studies (for example Gächter et al. (2013)) have used this.

differences were found in the mean bribe offered. Finally, there was no framing effect found in the public official's decision to accept or reject bribes. To summarize, framing effects have been found in some cases while not in others.<sup>8</sup>

Barr and Serra (2009) claim that their absence of framing effects<sup>9</sup> in the role of the “public servant” is “consistent with the petty corruption frame appearing artificial to student subjects”. One wonders why students, who can identify themselves as bribe offering private citizens, cannot identify with bribe receiving public officials. Further, the conclusion of Cooper et al. (1999) that familiarity with the “underlying structure” of the game is important in order to find effects may hold true even without the subject population having real life experience with the exact roles which are being examined in the game. Therefore, the literature on framing in corruption games demands a further investigation on why framing effect may or may not be found.

This paper argues that the reasons for the negative result of Abbink and Hennig-Schmidt (2006) and mixed results found in Barr and Serra (2009) lie elsewhere. In order to find the effects of “psychological and social factors” the ideal control treatment is not merely a neutrally framed environment but one in which the expectations of the agents are suitably changed. The use of neutral language does not alter the psychological reference point as subjects by and large identify with the underlying structure, thereby resulting in no evidence of framing effects. The ideal counterfactual should in fact be one in which the reference frame itself is altered keeping the strategic choices identical. This is ensured not just by using a neutral language but also by altering the sense of entitlement which in turn helps change the reference frame. We argue that this is crucial in any study which seeks to investigate the role of a particular frame.

### 3 Experimental design

We implement a harassment bribery game<sup>10</sup> with a real effort task. Real effort task has been found to induce a sense of “ownership” (see for instance Hoffman et al. (1994); Ruffle (1998))—this is crucial for the sense of harassment to be triggered. Successful completion of the real effort task in the bribery game treatment leads to a prize for a citizen but a public official may demand a bribe in order to let her have the prize. Of course the citizen may accept or reject the demand depending on whether the value of the prize outweighs the (material and moral) cost of bribe. We then exploit this structure of the harassment bribery game and posit it as an ultimatum game where successful completion of the real effort task merely leads to qualification for Player A to play the next stage of the game. In the next stage Player B splits an amount (equivalent to the prize) between himself and Player A and the

<sup>8</sup> The lack of framing effects has also been found in dictator games (Dreber et al. 2013).

<sup>9</sup> Barr and Serra (2009) does find that the share of public servants who refuse to take a bribe is consistently higher in the corruption frame than in the abstract frame but the differences are not significant.

<sup>10</sup> Abbink et al. (2014) implements a version of harassment bribery game but it was not known to the author when this experiment was conceived.



latter decides whether to accept it or not. It is now clear that despite the change in the language of description and the sense of entitlement, the two games remain strategically identical. To examine whether the treatment effect comes from the change in language or the change in entitlement, we run a third treatment which implements the bribery game but in a neutral language, while keeping the nature of entitlement unchanged. This treatment is similar in spirit to the neutral language frames implemented in Abbink and Hennig-Schmidt (2006) and Barr and Serra (2009).<sup>11</sup> Finally, ours is an entirely pen and paper experiment.

### 3.1 First experiment

#### 3.1.1 Harassment bribery game treatment

Figure 1a lays out the bribery game (BG henceforth). A citizen (C) has fifteen minutes to perform a real effort task.<sup>12</sup> The task is graded by a matched public official (PO) who is seated in another room. If a citizen scores 10 or above she “passes” the test and is entitled to a prize of 400 *Mohars* (M400). If she fails to obtain at least 10 then she earns only the participation fee of M200.

However, even if the citizen scores 10 or above, the public official may demand a bribe of amount  $b$  in order to let the citizen have her prize. Notice that the bribe of amount  $b$  is an extract from the value of the citizen’s entitlement, which in this case is M400. Any  $b > 400$  is likely to be rejected by a reasonable citizen. So, we let  $b \in \{0, 20, 40, \dots, 400\}$ . Upon receiving a bribe demand for  $b$ , a citizen may accept it or reject it. If she accepts the demand for a bribe, she earns  $M(200 + 400 - b)$  whereas if she rejects it, she earns the participation fee of M200 only.

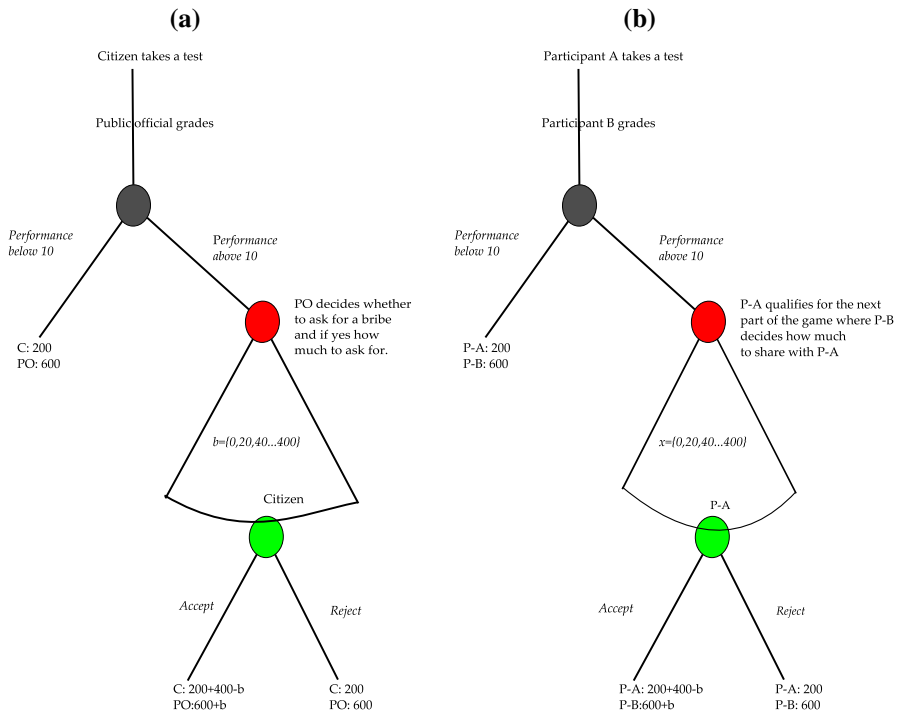
Public Official on the other hand gets a participation fee of M200 and a salary of M400 for the task of approving the citizens. Thus, if he demands a bribe  $b$  and his demand is accepted, then his earnings are  $M(200 + 400 + b)$ . If his demand is rejected however, his earnings are  $M(200 + 400)$  i.e. M600 only. Since the official earns at least as much as the citizen, an act of bribery cannot be explained by alternative explanations such as inequity aversion. This is a one shot game which reflects the standard harassment bribe situation where a briber and a bribee meet only once.

Since our interest is in emulating harassment bribe situations, the real effort task which a citizen undertakes is carefully calibrated such that she is more likely to

<sup>11</sup> There is a fourth possibility of a complete  $2 \times 2$  design—one where the ultimatum game treatment is framed in terms of loaded language but passing the test only qualifies one to the next stage of the game. Since a citizen does not deserve a prize, the question of bribe becomes irrelevant. One can however introduce an ultimatum game at the second stage with a neutral word for bribe such as transfer. With a mix of loaded terms (citizen, public official) and neutral terms (transfer instead of bribe), it is not clear what the treatment will measure.

<sup>12</sup> A version of the matrix or box task, introduced by Mazar and Ariely (2006), was used. In the task subjects had fifteen minutes to find two numbers in a 3 by 3 matrix which added up to 10. For example a matrix may have 4.55 and 5.45 which adds up to 10. There were twenty such matrices. The task was specifically chosen to ensure that subjects knew how many they had solved correctly. Thus, they knew whether they had won the prize and any demand for a bribe might have been considered unfair.





**Fig. 1** Harassment bribery game (BG) and ultimatum game (UG). **a** BG, **b** UG

cross the threshold score of 10.<sup>13</sup> At the same time the task itself is crucial as it induces a sense of entitlement and hence a sense of harassment in her.

### 3.1.2 Ultimatum game treatment

Figure 1b lays out the ultimatum game treatment (UG henceforth). In this treatment Participant A (analogous to citizens in BG) performs the real effort task which in turn is graded by a randomly matched Participant B (analogous to public officials in BG). If Participant A (P-A henceforth) solves 10 or more boxes correctly within the stipulated fifteen minutes, then she *qualifies* for the next round of the game in which Participant B (P-B henceforth) decides to keep  $Mx$  for himself and make an offer  $M(400 - x)$  to P-A. Thus in this case P-A's performance in the test only determines whether she proceeds to the next part of the game or not.<sup>14</sup> If she does proceed to the next stage, then she can decide whether to accept or reject the offer in which cases she receives  $M(200 + 400 - x)$  and  $M200$ , respectively. If P-B's offer is accepted

<sup>13</sup> Notice that letting citizens pay a bribe when they have failed the test is bribery for sure but does not amount to harassment bribery. The ultimatum game parallel applies only for the latter and thus, we were specially interested in it—hence the calibration. In this treatment 95 % of the participants scored at least 10.

<sup>14</sup> Calibration of the task ensured that 92.5 % of the subjects qualified for the next round.

then P-B gets  $\mathbf{M}$  ( $200 + 400 + x$ ), if it is rejected, then he only gets  $\mathbf{M}$  ( $200 + 400$ ), i.e.  $\mathbf{M}$  600. In this way the treatment represents a strategically identical version of BG but there is a subtle change in the sense of entitlement.

This alternative frame is designed to change the sense of guilt or the psychological moral cost of indulging in a corrupt act.<sup>15</sup> The naming of subjects as “Participant A” and “Participant B” instead of “Public Official” and “Citizen” and using words like “transfer” instead of “bribe” are meant to induce value neutrality. Additionally, the expectations of the subjects and their psychological reference points, are now different owing to a change in the entitlement since UG triggers social norms about sharing which BG triggers social norms of not taking a bribe. A comparison of behavior in BG and UG can help us answer whether or not moral costs are at play in BG.

### 3.1.3 Bribery game treatment with neutral labels

Harassment bribery treatment and the Ultimatum Game treatment differ with each other in two ways—one, the former is presented in a loaded language while the latter is described in neutral terms and two, the sense of entitlement among Citizens and Participant As is different. Thus, any observed difference between BG and UG can be attributed to either of the two changes. In order to precisely pin down if the observed effect comes from language or entitlement, we run an additional treatment—Bribery Game with Neutral labels (BG-N).<sup>16</sup> Very similar in spirit to the neutral language frames studied by Abbink and Hennig-Schmidt (2006) and Barr and Serra (2009), BG-N lays out the harassment bribery game (Fig. 1a) in purely neutral language terms, without making any change in the nature of entitlement. As in UG described above, in BG-N too “Public Official” and “Citizen” are labeled as “Participant A” and “Participant B”, “bribe” is labeled as “transfer” and “demand” as “ask”. Participant As in BG-N are entitled to a prize of  $\mathbf{M}400$  upon passing the test. The treatment effect between BG and BG-N can thus be interpreted as the pure effect of change in language while that between BG-N and UG can be interpreted as the pure effect of the change in entitlement.

### 3.1.4 External validity

At the end of the treatments, the Acceptance/Rejection decisions are handed out in the BG and UG treatments, subjects in the role of P-Bs/public officials are asked to mentally compute their earnings in *Mohars* and then convert them into Rupees at a previously declared exchange rate. Several examples of how to calculate the earnings are provided at this point. Each subject then finds an envelope underneath

<sup>15</sup> A referee rightly points out that the UG frame not just eliminates the moral cost of indulging in corruption but replaces it with the moral cost associated with ultimatum games. It is governed by its own moral frame driven by a set of non-monetary motivations which are distinctly different from the one in the BG frame.

<sup>16</sup> I thank two anonymous referees for suggesting this treatment.

his desk, consisting of Rs. 10 currency notes.<sup>17</sup> Subjects are asked to take out the money that they have earned and leave the rest in the envelope. After these instructions, the experimenters leave the room. The subjects then take their earnings and leave the room one by one.<sup>18</sup>

Subsequently we match the amount of money left in the unmarked envelopes with the response sheets and exit surveys. This allows us to see if a subject has taken more money than his earnings. The objective of this exercise is to identify two behavioral correlates—first, between the decision to seek a bribe and the decision to steal more money from the envelope. Second, between the amount demanded as bribe and the amount of money stolen. Thus in this setting corruption behavior in a framed lab behavior can be juxtaposed with stealing behavior.

### 3.1.5 Experimental procedure

The experiments were conducted in India, with student subjects recruited from Jadavpur University and Presidency University in Kolkata for BG and UG and those recruited from Bangalore for BG-N. A total of three hundred and thirty students were recruited and each subject participated in only one role of one treatment.

Each session consisted of about twenty subjects with half in each role. The subjects were asked to report to two different rooms physically separated by a fair distance in order to maintain anonymity. The instructions for the respective roles were read out (See Appendix 3). Several examples were worked out and earnings of each role were spelt out both in *Mohars* and Rupees. Answer sheets from the Citizens or Participant As were then handed out to the matched Public Officials or Participant Bs who graded them with the help of a solution manual provided to them. While the Citizens/P-As filled out the short exit survey, their answer sheets were graded by the POs/P-Bs. After grading, they wrote their decisions about bribe or transfer in the response sheets. These response sheets were subsequently returned to the other room for acceptance or rejection by Citizens/Participant As. Following this, the Public Officials and P-Bs filled out their exit survey.

The Citizens, the Player As were paid in cash. The Public Officials or Player Bs in BG and UG were asked to calculate their earnings after letting them have the response sheets back. Then they were asked not to communicate among themselves and to take their earnings from the envelopes which were placed underneath their desks, following which the experimenters left the room. Participant Bs in BG-N were paid in cash at the end of the experiment.

The sessions lasted for approximately 1 h. The instructions were read out in English and then paraphrased in Hindi and Bengali. All payoffs for this experiment were stated

<sup>17</sup> The information that an envelope filled with money is placed underneath the desk is revealed only at the end when they are about to receive their earnings. So, this does not have any impact on the earlier behavior.

<sup>18</sup> One referee points out that this in the true sense is not equivalent to observing the subjects in the field. Though they were not observed by the experimenter, they were still in a lab. We acknowledge that this is not the best design to capture external validity of unethical behavior and perhaps that is why the results are, at best, weak. However, subjects felt unobserved and it was not obvious to them that the remaining number of notes in the unmarked envelope would be matched back to their responses.

in terms of a fictitious currency called *Mohar*.<sup>19</sup> At the end of the experiment participants exchanged their earnings for Rupees at the rate 100 *Mohars* = Rs. 50. The earnings ranged from Rs. 100 to 450 with an average of Rs. 203 (~6 USD in Purchasing Power Parity terms).

### 3.2 Second experiment

#### 3.2.1 Design

This experiment elicits social appropriateness of the UG, BG, BG-N treatments using a tool, developed by Krupka and Weber (2012). The subjects are rewarded if their appropriateness ratings match the ratings given by most other people in the room. The reward, attached to the mode of the distribution of ratings, helps elicit a second order belief about the societal (defined as the others in the classroom) viewpoint. We interpret this as the social norm governing the situation.<sup>20</sup>

Subjects are shown either the situation in the BG treatment or UG treatment or BG-N treatment. Then they were asked to rate each action for each role as one of the following—“very socially inappropriate”, “socially inappropriate”, “somewhat socially inappropriate”, “value neutral”, “somewhat socially appropriate”, “socially appropriate” and “very socially appropriate”. The responses were later converted into numerical scores of −3, −2, −1, 0, +1, +2 and +3 respectively.

Each subject is paid a participation fee of Rs. 110. The action space of the public official/P-B is divided into several categories depending on the level of bribe.<sup>21</sup> In the BG situation subjects are asked to report their social appropriateness rating for the actions of the public official as well as for actions of the citizen in regard to accepting or rejecting the bribe demand for all possible bribe categories. One category from the public official’s strategy space and one from the citizen’s strategy space are randomly selected. The modal response of the appropriateness rating for the randomly selected categories is noted. If a subject’s appropriateness rating for the randomly selected category for either the public official or the citizen is the same as the modal response, then she is paid double the participation fee, i.e. Rs. 220. The descriptions of the UG or BG-N situations are identical except that subjects are asked to rate the actions of P-A and P-B.

For the second experiment, a different set of subjects were invited to a classroom. A session described one of the three frames. Each session comprised of 20–30 subjects and four such sessions were conducted. A total of 40 subjects participated in the norm elicitation of BG and UG each, while 30 subjects participated in that of BG-N. We implemented an entirely between subject design in order to avoid potential experimenter demand effect and confounds arising from subjective assessments of the situations i.e. none of the subjects who participated in the UG and BG treatments, took part in the social norm treatment and none of the subjects rated more than one situation in the social norm experiment. The instructions were read out in English and then paraphrased in Hindi and Bengali.

<sup>19</sup> *Mohar* or gold coin was a precious unit of exchange used in medieval India, an etymological history of which can be traced back to the Persian word *mohr*, meaning seal.

<sup>20</sup> While both the situations in BG and UG are governed by injunctive norms, i.e. what one should or should not do, the nature injunction that works for the two frames are very different.

<sup>21</sup> Figure 5 and Table 4 in Appendix 1 lays out the categories.

## 4 Results

Since we were interested in the phenomenon of harassment bribery, the general design of the experiment and the task were so calibrated that the majority of the subjects crossed the threshold. Three subjects in the UG and two in the BG treatment scored less than 10. However, 38 pairs of subjects in each treatment fulfilled this criteria as one P-A was wrongly classified as having crossed the threshold.<sup>22</sup> The analysis of BG-N is performed with data from 30 pairs of observations. In the following discussion (but not in the actual experiment) we use the word *ultim* to denote the UG treatment equivalent of the bribe i.e. the amount that P-B proposes to keep for himself.

### 4.1 Framing effects (experiment 1)

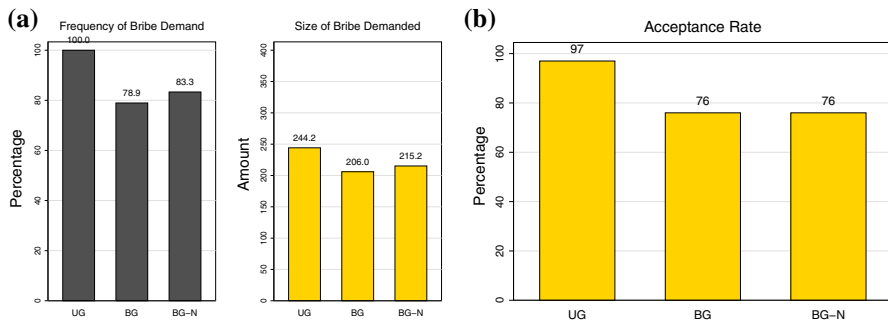
We find significant differences across BG and UG treatments both in terms of the proportion of subjects who demanded a bribe/*ultim* and the amount of bribe demanded. We also find important differences in the acceptance rate between the two treatments. As Fig. 2a shows, all the subjects playing the role of Participant B in UG propose a split with a positive amount for themselves while only 78.9 % of the Public Officials demand a bribe in BG. Table 1 reports the  $p$  values, clustered over sessions and rejects the null of equality of proportions ( $\chi^2$  test,  $p$  value = 0.07). The mean *ultim* proposed was M244.2 and this was significantly higher than the mean bribe demanded which was M206 (Mann–Whitney test,  $p$  value = 0.08). There was however no statistical difference in either the frequency of bribe demand between the BG and BG-N treatments ( $\chi^2$  test,  $p$  value = 0.65) or the bribe amount between them (Mann–Whitney test,  $p$  value = 1.00).<sup>23</sup> Thus, the difference in behavior in BG and UG does not arise from the difference in the language used but arises from the difference in the sense of entitlement in the two frames. This result is in line with results obtained in the past and suggests that neutral language in itself may not be enough to impose an alternative reference frame.

In Fig. 2b we investigate the acceptance rate of the Citizens and Participant Bs. While 97 % accepted the division proposed in the UG treatment, only 76 % accepted to pay the bribe (if one was demanded) in BG and BG-N. The difference is statistically significant ( $\chi^2$  test, clustered  $p$  value = 0.03) as reported in Table 1. Furthermore, though the average *ultim* proposed was larger than the average bribe demanded, only one offer was rejected in UG (where *ultim* proposed was 400); on the other hand bribes which were rejected had a mean of M237. Thus, when posited in terms of bribe much lower amounts were rejected in the BG treatment.<sup>24</sup>

<sup>22</sup> Presumably because the P-B realized that he could increase his own earnings by letting the P-A pass.

<sup>23</sup> If zeroes were included then the average bribe demanded in BG was M163 while that in BG-N was M179 (MW-test,  $p$  value = 0.76)

<sup>24</sup> Lack of sufficient number of observations make mean comparison test redundant.



**Fig. 2** Treatment differences. **a** Frequency of the bribe demand (bribe >0): The difference between UG and BG is statistically significant ( $\chi^2$  test,  $p$  value = 0.06) but that between BG and BG-N is not statistically significant. Size of the bribe demand (size of bribelbribe >0): Difference between UG and BG is statistically significant (Mann–Whitney test,  $p$  value = 0.08) but that between BG and BG-N is not (Mann–Whitney test,  $p$  value = 1). **b** Acceptance rates conditioned on the ultim/bribe demanded. The numbers are stated in terms of ultim/bribe. The difference in acceptance rates between UG and BG is statistically significant ( $\chi^2$  test,  $p$  value = 0.03) but that between BG and BG-N is not ( $\chi^2$  test,  $p$  value = 0.97).

We shall explain the difference in the treatment effects in terms of the difference in the fairness perception of the two treatments in Sect. 4.4.

## 4.2 Regression results

The main result from the regression analysis, laid out in Table 2, is consistent with the result from the mean difference tests discussed above—the bribe amount in the BG treatment is M42 lower than the *ultim* demanded in the UG treatment even after controlling for gender, age, caste and log of household income [for a study of how cheating behavior depends on demographic characteristics, see Bucciol et al. (2013)]. The difference is statistically significant as indicated in specification (2). Specification (4) and (5) uses the full sample data, conditioned on positive bribe demand, and regresses bribe amount on the BG-N and UG treatment dummy.<sup>25</sup> The BG-N treatment dummy is not significantly different from zero while the UG treatment dummy is. This further suggests that the effect of BG-N is statistically indistinguishable from that of BG while in UG the bribe demand is significantly higher than in BG, even after including control variables. Furthermore, the marginal effects from the estimated probit model for acceptance of a bribe/*ultim* in (6)–(7) reveal that the probability of accepting to pay a bribe is 0.27–0.28 lower than that of accepting to share an *ultim*. Specification (8) uses the full sample data to estimate how BG-N and UG treatment dummy affects the acceptance probability. Once again the results show that the coefficient of BG-N is insignificant while that of UG is significant. The evidences from the regression analyses thus suggest that subjects behave very similarly irrespective of whether they operate under a loaded frame or a

<sup>25</sup> We thank an anonymous referee for suggesting this specification.

**Table 1** Mean difference

	Mean BG	Mean UG	Mean BG-N	Difference (1)	Difference (2)	Difference (3)
PO/B's decision to ask for a bribe/ <i>ultim</i>	0.790	1.000	0.83	0.21***	0.04	0.17**
N	38	38	30			
<i>p</i> value ( $\chi^2$ test)				0.003	0.65	0.01
<i>p</i> value (clustered $\chi^2$ test)				0.067		
C/A's decision to accept the offer	0.763	0.975	0.766	0.211**	0.003	0.21**
N	38	38	30			
<i>p</i> value ( $\chi^2$ test)				0.007	0.973	0.01
<i>p</i> value (clustered $\chi^2$ test)				0.034		
Bribe/transfer amount (full sample)	162.6	244.2	179.33	77***	16.7	64.88***
N	38	38	30			
<i>p</i> value <i>t</i> test				0.000	0.53	0.001
<i>p</i> value clustered <i>t</i> test				0.001		
<i>p</i> value rank sum tests				0.002	0.76	0.002
Bribe/transfer amount (restricted sample)	206	244.2	215.2	38.2	9.2	29.01***
N	30	38	25			
<i>p</i> value <i>t</i> test				0.023	0.67	0.056
<i>p</i> value clustered <i>t</i> test				0.135		
<i>p</i> value rank sum tests				0.084	1.00	0.034

\*\*\* denotes significance at the 1 % level, \*\* at the 5 % level and \* at the 10 % level from clustered *t* test. Difference (1) reports and tests the difference in mean between BG and UG; difference (2) between BG and BG-N; difference (3) between BG-N and UG. Clustering was done for each session. Treatment BG-N was conducted in one session, hence only one cluster. The acceptance decision reported above is from the entire sample i.e. it is not conditioned on being asked for a bribe/*ultim*. The restricted sample consists of only those who demanded a bribe



**Table 2** Regression Results

	Amount demanded				Acceptance		
	OLS				Probit		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
BG Treatment	-38.2*** (16.38)	-42.46** (18.26)				-0.27*** (0.46)	-0.28*** (0.09)
BG-N Treatment			-29.01* (14.88)	9.20 (18.35)	31.16 (23.6)		0.07 (0.05)
UG Treatment				38.21** (16.55)	41.64** (17.47)		0.26*** (0.08)
Gender		-34.06* (18.26)			-34.60** (14.99)		0.03 (0.05)
Age		-1.04 (3.93)			-1.07 (3.13)		-0.02 (0.02)
Caste		-5.90 (28.84)			1.30 (18.38)		-0.04 (0.04)
Log of household income		-8.80 (7.03)			-7.00 (9.42)		-0.07 (0.04)
Constant	244.2 (5.46)	-370.16** (133.10)	244.21** (9.37)	206*** (12.37)	308.86** (123.58)		
Amount demanded						-	-0.001* (0.001)

Table 2 continued

	Amount demanded				Acceptance			
	OLS				Probit			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
N	68	68	63	93	93	68	68	93
R <sup>2</sup>	0.08	0.85	0.06	0.06	0.13	0.19	0.26	0.39

\*\*\* denotes significance at the 1 % level, \*\* at the 5 % level and \* at the 10 % level. Specification (4), (5) and (8) uses the full sample data to estimate the coefficients of BG-N and UG treatment. The reported  $R^2$  for probit models is Pseudo  $R^2$ . The BG/BG-N/UG Treatment takes value 1 for BG/BG-N/UG and 0 for others. Model (1) and (3) reports the mean difference between BG and UG and BG-N and UG, respectively. The variable caste takes value 0 for upper caste and 1 for backward castes. Column (6), (7) and (8) report the marginal effects from probit estimations

neutral one; the difference in behavior clearly comes from a difference in entitlement.

### 4.3 External validity

We report the correlations between stealing and bribe demand decisions only in BG.<sup>26</sup> Findings, reported in Table 3, are inconclusive but indicate a (weak) positive correlation between bribe demand and stealing. The correlation between the binary choice decision to ask for a bribe and steal is 0.22 and that between bribe amount and stolen amount is 0.3. While statistically insignificant, it is similar to the correlations reported in other studies which look at external validity. For example, Baran et al. (2010) report correlation of 0.29 between field setting and lab while Englmaier and Gebhardt (2011) and Franzen and Pointner (2013) report 0.4 and 0.2, respectively. In fact Franzen and Pointner (2013) conclude by saying that despite the weak or moderate association, their results should still be viewed as the glass half full (i.e. lab and field behaviors are more or less similar).

A further probe reveals that there is a strong negative correlation between stealing and acceptance to pay a bribe (Spearman Corr. Coeff =  $-0.61$ ,  $p$  value = 0.00). This indicates that we picked up an effect where the probability of stealing increases with the refusal on the part of the citizen to pay a bribe, i.e. unethical behavior at one level crowds out that in another. We interpret these correlations as that the lab corruption game is most likely appealing to the same unethical behavior among subjects as the stealing<sup>27</sup>, with an important caveat that correlations obtained here should be probed further before conclusive claims can be made.

### 4.4 Social norms (experiment 2)

Our discussion above in Sect. 3 conjectured that the two treatments—BG and UG, though strategically identical, were governed by different perceptions of social appropriateness. The social appropriateness ratings of UG and BG were elicited from 40 subjects each while those of BG-N were elicited from another 30. Figure 4 in Appendix 1 gives a visual description of how the distribution of norm rating changes across the three treatments. Table 4 lays out the full distribution of the ratings corresponding to each action of Participant B/Public Official in UG, BG-N and BG, respectively. It also presents the mean ratings for both treatments and the  $p$  values reported by the Mann–Whitney ranksum test for the difference in mean ratings.

<sup>26</sup> Though the Participant Bs in UG had the opportunity to steal, it is not clear how to interpret the correlation between *ultim* demand and stealing. Hence we don't report it.

<sup>27</sup> Direct stealing of cash and deliberately failing to return excess money are two alternative ways in which unethical behavior can be examined on the field however one may note that these two methods appeal to different degrees of unethicality. Our conjecture is that the latter will result in fewer observations in the false-positive domain. Also note that none of our other results hold when stealing is used as the dependent variable.

**Table 3** External validity

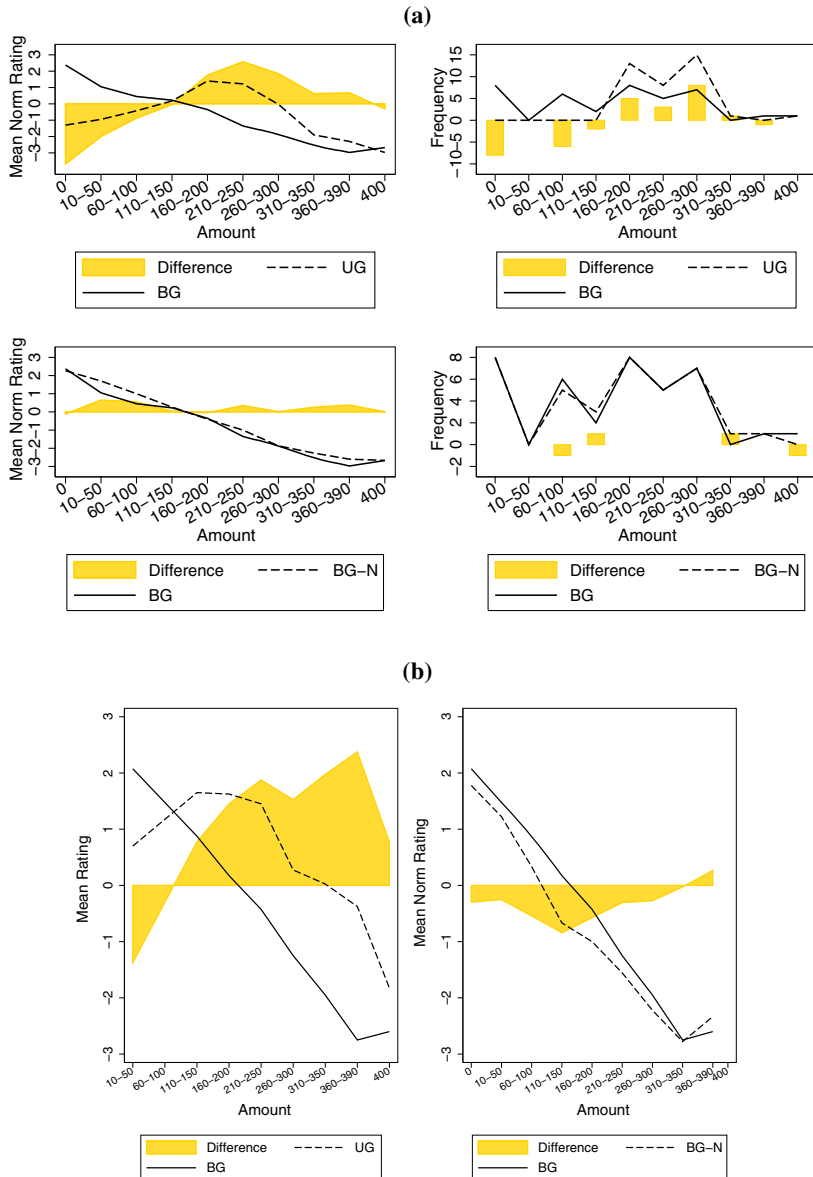
Variable 1	Variable 2	Correlation coefficient	<i>p</i> value
1(bribe > 0)	1(Stealing > 0)	0.22	0.177
Successful bribe extraction	1(Stealing > 0)	−0.61***	0.000
Bribe amount   bribe > 0	Stolen amount	0.30	0.150

\*\*\* denotes significance at the 1 % level, \*\* at 5 % level and \* at 10 % level. Spearman correlation coefficient for binary variables. 1(.) denotes identity variable

While in the following analysis we focus on the mean social appropriateness ratings, replacing the mean with median or mode does not make a difference to the results. The results show that for low levels, bribe demanded is considered socially more appropriate than comparable amount of *ultim*. However, for medium and higher levels, bribe is seen to be socially more inappropriate than *ultim*. This indicates that social sanction against a small bribe amount is nearly absent. In the UG treatment an equal split ranks high on the social appropriateness scale and is significantly higher than that of a bribe demand of M200 (MW test, *p* value = 0.00). No statistical difference is observed for higher levels of *ultims* or bribes however. On the contrary BG-N and BG are governed by very similar social appropriateness norm.

The top row in Fig. 3a compares the social norm and actual behavior of UG and BG, while the figures at the bottom compare the same between BG-N and BG. While the left hand side panels show norm ratings for the two treatments and their difference, the right hand panels shows the distribution of actual behavior. The top row confirms the co-movement between the difference in the norm ratings between UG and BG, demonstrated by the area plot in gold on the left and the difference in actual behavior given by the gold bars. The non-parametric correlation between the difference in the mean norm rating and the difference in actual behavior is positive and significant (Spearman Corr. Coeff = 0.78, *p* value = 0.01), reaffirming the co-movement discussed above. The point that actual behavior in BG and UG is governed by social appropriateness norm of the frames is further illustrated by the fact that when number of observations in each category (see Fig. 4 in Appendix 1 for the categories and the mean norms in each category) is regressed on the mean social norm rating for each category, the regression coefficient is 1.51 (*p* value < 0.001).<sup>28</sup> Thus, despite the between subject design in eliciting social appropriateness, we do find clear evidence that the frequency of observation in each category increases with social appropriateness, indicating that views of how appropriate an action is, dictate how many or how few people go for it. This effect is, in some ways, a weaker version of earlier evidence of contagion effects as found in Innes and Mitra (2013). While contagion

<sup>28</sup> This regression is done with only 20 observations—however, that the standard assumptions of ordinary least square regressions still hold is evident from the diagnostics of Fig. 5 included in the Appendix 2. The result of an outlier test based on inter-quartile range shows the presence of only one outlier. Re-estimating this specification by excluding the outlier does not lead to any significant difference.



**Fig. 3** Norm and actual behavior. **a** Comparing social norms with amount demanded in UG, BG and BG-N. **a** The *top two panels* compare the norm ratings and actual behavior in UG and BG. The *bottom two panels* compare the same in BG-N and BG. In the *left hand side panels* the y-axis represents the range of possible norm ratings, i.e. -3 represents very socially inappropriate and 3 represents very socially appropriate. The *shaded area* represents the difference. The *right hand side panels* compare the actual frequency of choices for equivalent amounts. The *shaded bars* represent the difference in frequencies. Clearly social norm and actual behavior are very similar in BG-N and BG whereas they are very different in UG and BG. **b** Social norm of acceptance in UG, BG and BG-N. **b** The *left (right) hand side panel* compares the social norm ratings of acceptance of demand between UG and BG (BG-N and BG). The *shaded area* represents the difference. Social norm of acceptance is similar in BG-N and BG but it is very different in UG and BG

effects demonstrate herd behavior and hint at diminishing marginal cost at work by following others, our results show that there is indeed a set of normative codes governing our actions. This works quite independently of the information about what others do.

The bottom row of Fig. 3a compares the social appropriateness rating and actual behavior in BG-N and BG. Clearly, the two frames are governed by very similar social appropriateness norms indicating once again that a change to neutral language by itself does not change the moral environment governing the two frames—a possible reason why we do not observe a statistical difference in behavior between the two frames in the first place.<sup>29</sup>

The social appropriateness rating of accepting the proposed split in UG and BG is shown in the left panel of Fig. 3b. Accepting to pay a small bribe is considered socially appropriate but the inappropriateness steadily increases with the bribe amount. For the UG treatment the appropriateness rating for accepting an equal split is highest and then it declines though it remains significantly higher than that of the BG treatment for equal amounts. The left hand panel compares the social appropriateness of acceptance in BG-N and BG. Once again it is clear that language by itself makes little difference to the moral environment governing a decision.

To summarize, BG and UG frames elicit different perceptions of appropriateness—equal sharing is considered very appropriate in UG but demanding no or a small amount of bribe is considered appropriate in BG, though the two games are strategically identical. Once again the difference in social appropriateness norms between BG and UG can be traced to a difference in entitlement but not to a difference in language. It is also worth noting that the fact that small bribes elicit high appropriateness measures implies that there is little social sanction associated with smaller levels of corruption in India where the experiment is conducted. This may give us a clue about why lack of social sanction against petty bribery helps it perpetuate in other similarly placed countries.

## 5 Discussion and conclusion

The canonical structure of a laboratory corruption game essentially implements splitting of a surplus which is generated when those in positions of power (e.g. public officials) unlawfully favor those in need of services (e.g. firms or citizens), for instance look at Abbink et al. (2002), Cameron et al. (2009), Barr and Serra (2010) and more recently Gneezy et al. (2013). To this skeletal, parameters are then altered, layers added and context changed. The harassment bribery game, which we implement in this paper, remains true to this canonical structure where a surplus is

<sup>29</sup> Qualitative interactions after the BG-N treatment revealed that the subjects thought it was Participant A's prize and it was unfair for Participant B to take away a share of her prize.

split. We compare the behavior of subjects in the harassment bribery game and a strategically identical but differently framed ultimatum game. The bribery and the ultimatum game frames, though identical in strategic terms, differ in two respects—one, the former is described in terms of emotionally loaded language while the latter is described in neutral language and two, the frames differ in terms of the psychological reference points of the subjects, which are induced through a difference in the sense of entitlements. In order to precisely identify which of the two is responsible for the observed difference in behavior, we devise a third treatment which implements the bribery game in neutral language, while keeping the nature of the entitlement identical.

Our results show that the two different moral frames trigger different responses among individuals. Not only do less people indulge in bribery, the average bribe demanded turns out to be significantly lower in the bribery game treatment than the average amount proposed for oneself in an ultimatum game. Furthermore, the patterns of acceptance to pay the bribe and to accept the split proposal vary widely as well. This points to the fact that moral costs are indeed at work when subjects operate under a corruption game. Furthermore, we find no treatment effect between bribery game in loaded language and that in neutral language indicating that altering the language in itself does not change the frame of reference. This may be one reason why past studies attempting to find framing effect in corruption games by solely changing to the neutral language, did not find one.

We then explain the behavioral differences observed in the two frames in terms of a measure of social appropriateness norm associated with each action in the two frames. Social appropriateness norm or commonly held perception about what is morally the right thing to do, successfully explains the pattern of the actual choices made in the two frames. Also, according to elicited social appropriateness, the bribery game is considered more inappropriate than the counterfactual.

In the light of the fact that past studies have not found decisive evidence of moral costs at play in laboratory corruption games, we also compare social appropriateness norm between the loaded and neutrally framed bribery games, as used in the past by Abbink and Hennig-Schmidt (2006) and Barr and Serra (2009). The fact that we fail to detect a difference in the social appropriateness norm between loaded and neutrally framed bribery games, further explains why framing effects had not been consistently found in the past—it may be that loaded and neutrally framed bribery games are not viewed in different lights by the subjects. In this paper we alter the psychological reference point of the subjects across the bribery game treatment and its counterfactual through a subtle change in the sense of entitlement, while keeping the two treatments otherwise strategically identical. We do find evidence of treatment effects indicating that moral costs indeed are at play in the bribery game. We then argue that in order to see evidence of moral cost at work, the appropriate



counterfactual is one which uses either loaded or neutral language but suitably alters the reference point of the subjects. It is the latter which helps create the alternative moral framework.

Having said that, we recognize that there are other differences between our set up and those of Abbink and Hennig-Schmidt (2006) and Barr and Serra (2009), which can potentially affect the results. For instance they conduct their experiments in western societies where corruption norms are very different from those in India where our experiment was conducted. Although the game we implement is sufficiently generic and true to the fundamental canonical structure of corruption games, it is not identical to the one which they implement. We implement a two person harassment bribery game unlike theirs which is a collusive bribery design. Barr and Serra (2009) in fact has a third person who is affected by a corrupt transaction. Despite the differences, we are tempted to believe that the hook of our results comes from a successful creation of an alternative moral framework, induced by the difference in the sense of entitlement. Interestingly, entitlement effects have also been observed in other strands of the literature, for instance Hoffman et al. (1994), Oxoby and Spraggon (2008) and Ruffle (1998) in dictator games and Hoffman et al. (1996) in ultimatum games. However, just like in our study, merely changing the frame by a different language has been found to have no effect in dictator games Dreber et al. (2013). Hence, our conclusion that in order to see a framing effect in corruption games, not merely the language but the very psychological reference point should be altered, holds true elsewhere in the literature as well.

The successful creation of an alternative moral framework is also indicated by the different societal perceptions of what is morally the right thing to do in the two environments. A clear evidence of moral cost at play along with evidence of moral indictment from society shows that laboratory corruption games do measure what they are set out to. This also ensures that this tool can be used to measure the change in unethical behavior by changing a relevant policy variable.

Our attempt to relate behavior in a lab with that in a field is met with limited success. Imaginative designs in the future may be able to do that more clearly. Furthermore, we find evidence of lack of social sanction against petty bribery. Besides institutional inadequacies, this may explain why the bad equilibrium of pervasive corruption is difficult to escape in many societies.

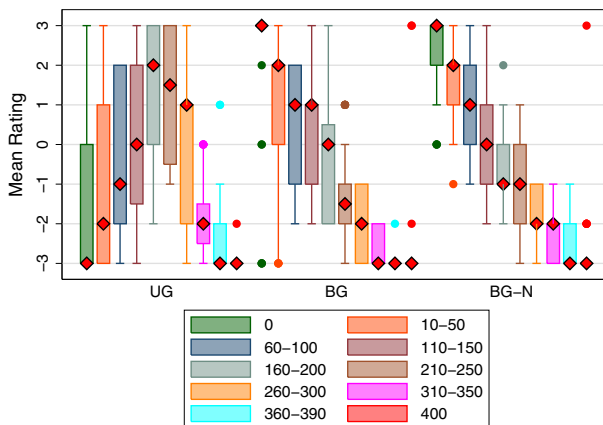
Finally, it is often argued, especially in the theoretical models which analyze corruption equilibrium in a rational actor framework, that our pursuit of ethicality is driven largely by self interest. Within the framework of this premise it becomes difficult to explain why the vast majority among us choose legitimate actions when a large number of dubious but materially better off alternatives are present. It seems unlikely that such a voluntary abdication of unethical behavior will be part of the resulting equilibrium without some norm induced behavior. One may conjecture that the complex web of laws and rules, which govern the modern society, may not be sustained if we are to act only with strategic selfishness vis-a-vis these laws. In

other words, our actions are disproportionately influenced by what we and others think is morally the right thing to do. Our experiment bears a testimony to that.

**Acknowledgments** I thank Goutam Gupta and Amitabha Chatterjee for letting me conduct the experiment at Jadavpur University and Presidency University, respectively. I am grateful to Abhishek Das for his excellent assistance while conducting the experiment. I am also thankful to Alexander Koch, John A List, Klaus Abbink, Roel van Veldhuizen, Arnab Mitra, Joydeep Bhattacharya, Marco Piovesan, Tor Eriksson, Bertil Tungodden, Alexander Cappelen and participants at TIBER XII and the 24th Jerusalem School in Economic Theory and two anonymous referees for their comments and suggestions. All remaining errors are mine. The grant for the experiment came from the Department of Economics and Business, Aarhus University.

## Appendix 1

See Fig. 4.



**Fig. 4** Distribution of norm rating for each category of *ultim*/bribe in UG and BG. The figure maps the entire distribution of the social norm ratings for UG, BG and BG-N treatments. For example, the *blue box* plot corresponds to the norm ratings for a bribe/ultim demand between 60 and 100 (from the legend). The social norm ratings for that amount of ultim (bribe) may be read from the *left (right) panel*. The *diamond marks* the median of the distribution of ratings for each category. The social norm of UG treatment is an inverted U with the equal split considered as most appropriate. The social norm of the BG and BG-N treatments decline monotonically indicating that social appropriateness goes down with the bribe amount, irrespective of framing. (Color figure online)

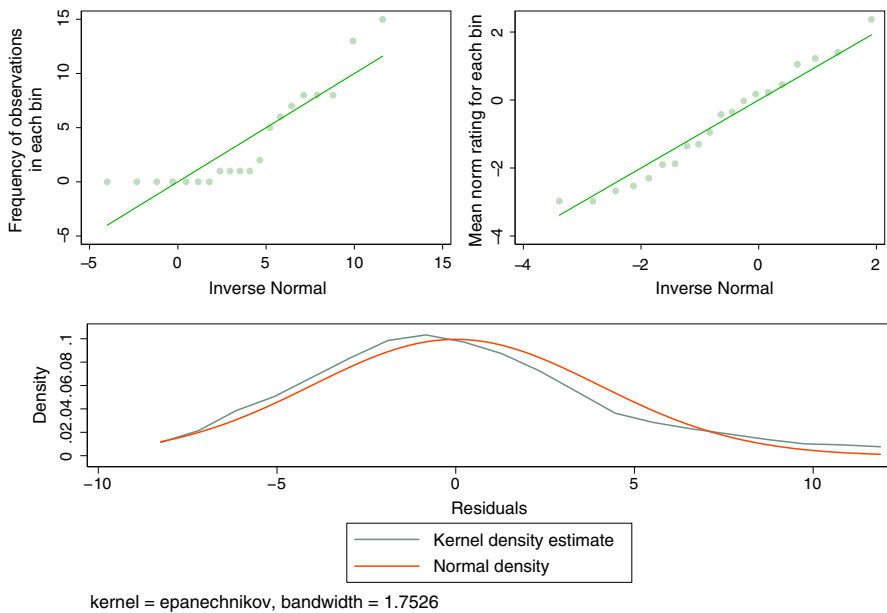
Table 4 Frequency distribution of norm ratings

Bribe/ <i>Ultim</i>	Mean UG							Mean BG						
	-3	-2	-1	0	1	2	3	-3	-2	-1	0	1	2	3
0	55	10	10	0	2.5	5	17.5	7.5	0	0	5	0	2.5	85
10-50	35	20	10	5	12.5	5	12.5	7.5	0	12.5	15	10	35	20
60-100	12.5	32.5	10	5	10	30	0	0	10	25	10	20	35	0
110-150	5	20	15	12.5	17.5	20	10	0	22.5	12.5	12.5	30	17.5	5
160-200	UG	0	5	2.5	25	10	30	BG	0	30	17.5	27.5	12.5	7.5
210-250		0	0	25	5	20	22		15	35	32.5	5	12.5	0
260-300		5	25	17	0	30	17.5		27.5	32.5	40	0	0	0
310-350		25	50	15	10	0	0		52.5	47.5	0	0	0	0
360-390		52.5	35	7.5	0	5	0		97.5	2.5	0	0	0	0
400		97.5	2.5	0	0	0	0		92.5	2.5	5	0	0	0

Bribe/ <i>Ultim</i>	Mean BG-N							p-value (2)	
	-3	-2	-1	0	1	2	3		
0	0	0	0	10	6.7	30	53.3	2.27	0.02
10-50	0	0	3.3	3.3	30	46.7	16.7	1.7	0.26
60-100	0	0	6.7	20	43.3	26.7	3.3	1	0.20
110-150	0	3.3	23.3	33.3	26.7	10	3.3	0.27	0.93
160-200	BG-N	0	6.7	53.3	16.7	20	3.3	0	0.86
210-250		3.3	30	36.7	23.3	6.7	0	0	-1
260-300		20	46.7	33.3	0	0	0	0	-1.87
310-350		43.3	40	16.7	0	0	0	0	-2.27
360-390		63.3	33.3	3.3	0	0	0	0	-2.6
400		83.3	13.3	3.3	0	0	0	0	-2.67

Note: *p*-value (1) and *p*-value (2) report *p*-values from the Mann Whitney Rank Sum test of the mean difference between ratings of UG and BG and BG and BG-N, respectively. Shaded cells for each category in each treatment denote the modal ratings.

## Appendix 2



**Fig. 5** A small sample regression, like the one described in Footnote 28, produces biased estimates. To the extent that the distribution of the data can help, the quantile plots of the frequency of observations in each bin and the mean of social norms rating for each bin are given here. The distribution of residuals from the regression is also included. The result of the outlier test, based on inter-quartile range, shows the presence of one outlier. Re-estimating the specification and excluding the outlier does not lead to any significant difference.

## References

- Abbink, K., & Hennig-Schmidt, H. (2006). Neutral versus loaded instructions in a bribery experiment. *Experimental Economics*, 9(2), 103–121.
- Abbink, K., Irlenbusch, B., & Renner, E. (2002). An experimental bribery game. *Journal of Law, Economics and Organization*, 18(2), 428–454.
- Abbink, K., Dasgupta, U., Gangadharan, L., & Jain, T. (2014). Letting the briber go free: An experiment on mitigating harassment bribes. *Journal of Public Economics*, 111(C), 17–28.
- Alatas, V., Cameron, L., Chaudhuri, A., Erkal, N., & Gangadharan, L. (2009). Subject pool effects in a corruption experiment: A comparison of Indonesian public servants and Indonesian students. *Experimental Economics*, 12(1), 113–132.
- Armanter, O., & Boly, A. (2013). Comparing corruption in the laboratory and in the field in Burkina Faso and in Canada. *The Economic Journal*, 123, 1168–1187.
- Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3), 797–817.
- Banerjee, R., Baul, T., & Rosenblat, T. (2015). On self selection of the corrupt into the public sector. *Economics Letters*, 127, 43–46.
- Banuri, S., & Eckel, C. (2012). Experiments in culture and corruption: A review. In W. D. S. Leonard (Ed.), *New advances in experimental research on corruption, research in experimental economics, chapter 3* (pp. 51–76). Bingley: Emerald Group Publishing Limited.

- Banuri, S., & Eckel, C. (2012). On the effects of culture on punishment of bribery: US versus Pakistan, CBEES Working Paper Series 11-05.
- Baran, N. M., Sapienza, P., & Zingales, L. (2010). Can we infer social preferences from the lab? Evidence from the trust game. NBER Working Papers 15654, National Bureau of Economic Research, Inc January 2010.
- Barr, A., & Serra, D. (2009). The effects of externalities and framing on bribery in a petty corruption experiment. *Experimental Economics*, 12(4), 488–503.
- Barr, A., & Serra, D. (2010). Corruption and culture: An experimental analysis. *Journal of Public Economics*, 94(11–12), 862–869.
- Bernheim, B. D. (1994). A theory of conformity. *Journal of Political Economy*, 102(5), 841–877.
- Bicchieri, C., & Xiao, E. (2009). Do the right thing—but only if others do so. *Journal of Behavioral Decision Making*, 22, 191–208.
- Buccioli, A., Landini, F., & Piovesan, M. (2013). Unethical behavior in the field: Demographic characteristics and beliefs of the cheater. *Journal of Economic Behavior & Organization*, 93(C), 248–257.
- Camerer, C. (2007). The promise of lab-field generalizability in experimental economics: A reply to levitt and list. In F. Guillaume & A. Schotter (Eds.), *Methods of modern experimental economics*. Oxford: Oxford University Press.
- Cameron, L., Chaudhuri, A., Erkal, N., & Gangadharan, L. (2009). Propensities to engage in and punish corrupt behavior: Experimental evidence from Australia, India, Indonesia and Singapore. *Journal of Public Economics*, 93(7–8), 843–851.
- Cason, T. N., & Mui, V. -L. (1998). Social influence in the sequential dictator game. *Journal of Mathematical Psychology*, 42, 248–265.
- Cooper, D. J., Kagel, J. H., Lo, W., & Gu, Q. L. (1999). Gaming against managers in incentive systems: Experimental results with Chinese students and Chinese managers. *The American Economic Review*, 89(4), 781–804.
- Dreber, A., Ellingsen, T., Johannesson, M., & Rand, D. (2013). Do people care about social context? Framing effects in dictator games. *Experimental Economics*, 16(3), 349–371.
- Englmaier, F., & Gebhardt, G. (2011). Free-riding in the lab and in the field. Technical report 2011.
- Falk, A., & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, 326(5952), 535–538.
- Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, 8(4), 185–190.
- Fisman, R., & Miguel, E. (2007). Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets. *Journal of Political Economy*, 115(6), 1020–1048.
- Franzen, A., & Pointner, S. (2013). The external validity of giving in the dictator game. *Experimental Economics*, 16(2), 155–169.
- Gächter, S., Nosenzo, D., & Sefton, M. (2013). Peer effects in pro-social behavior: Social norms or social preferences? *Journal of the European Economic Association*, 11(3), 548–573.
- Gneezy, U., Saccardo, S., & van Veldhuizen, R. (2013). Bribery: Greed versus reciprocity. UC San Diego Working Papers, UC San Diego June 2013.
- Hoffman, E., McCabe, K. A., & Smith, V. L. (1996). On expectations and the monetary stakes in ultimatum games. *International Journal of Game Theory*, 25(3), 289–301.
- Hoffman, E., McCabe, K., Shachat, K., & Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior*, 7(3), 346–380.
- Robert, I., & Arnab, M. (2013). Is dishonesty contagious? *Economic Inquiry*, 51(1), 722–734.
- Kessler, J. & Vesterlund, L. (2014). The external validity of laboratory experiments: Qualitative rather than quantitative effects. In: F. Guillaume & A. Schotter (Eds.), *Methods of Modern Experimental Economics*. Oxford University Press, Oxford Forthcoming.
- Krupka, E., & Weber, R. A. (2009). The focusing and informational effects of norms on pro-social behavior. *Journal of Economic Psychology*, 30(3), 307–320.
- Krupka, E. L., & Weber, R. A. (2012). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of European Economic Association*, 11, 495–524.
- Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *The Journal of Economic Perspectives*, 21(2), 153–174.
- List, J. (2012). Using field experiments to understand the economics of crime, Working Paper 2012.
- Mazar, N., & Ariely, D. (2006). Dishonesty in everyday life and its policy implications, Technical Report 2006.

- Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives*, 14(3), 137–158.
- Oxoby, R. J., & Spraggon, J. (2008). Mine and yours: Property rights in dictator games. *Journal of Economic Behavior & Organization*, 65(3–4), 703–713.
- Reuben, E., & Riedl, A. (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior*, 77(1), 122–137.
- Ruffle, B. J. (1998). More is better, but fair is fair: Tipping in dictator and ultimatum games. *Games and Economic Behavior*, 23(2), 247–265.
- Serra, D. (2006). Empirical determinants of corruption: A sensitivity analysis. *Public Choice*, 126(1), 225–256.
- Serra, D., & Wantchekon, L. (2012). Experimental research on corruption: Introduction and overview. In W. D. S. Leonard (Ed.), *New advances in experimental research on corruption, research in experimental economics, chapter 1* (pp. 51–76). Bingley: Emerald Group Publishing Limited.
- Treisman, D. (2000). The causes of corruption: A cross-national study. *Journal of Public Economics*, 76(3), 399–457.