Power to the People? An Experimental Analysis of Bottom-Up Accountability of Third-Party Institutions

Published in: Journal of Law, Economics, and Organization, 2015, 31, 2, 347-382.
https://doi.org/10.1093/jleo/ewu007

Stefania Ottone

University of Milano-Bicocca and European University Institute

Ferruccio Ponzano

University of Eastern Piedmont

Luca Zarri*

University of Verona

This paper provides an experimental investigation of third parties' sanctioning behavior, in order to understand whether public officials (e.g., judges, politicians or regulators), when deciding about top-down interventions aimed at punishing wrongdoers, are sensitive to bottom-up pressure on the part of ordinary citizens, who are the major victims of wrongdoers' behavior. We set up a novel five-treatment design and compare situations where a wrongdoer acts under: 1) no third-party punishment; 2) nonaccountable third-party punishment and 3) accountable third-party punishment. We show that when citizens are active and make their voice heard, public officials sanction wrongdoing significantly more. Our experimental finding complements previous empirical work based on field data and suggests that when third-party institutions are held accountable, their propensity to fight misconduct is higher, other things equal. We view this result as good news with regard to domains where it implies that pro-consumer policies will be more likely (e.g. regulatory policies). The risk of pandering by elected officials and the danger of poorly informed decisions by the citizens are the flip side of the argument.

* Corresponding author. Email: luca.zarri@univr.it. We thank Alvin Klevorick, John Morgan and three anonymous referees whose comments helped us significantly improve the paper. Our gratitude also goes to Benedetto Gui and to the participants in the Urrutia Elejalde Summer School on Economics and Philosophy (Social Norms) in San Sebastian, the LabSi Workshops in Salerno and Florence, the IAREP/SABE Conference in Rome and the Young Economists' Workshop on Social Economy in Bologna (Forlì Center) for useful suggestions and comments. The usual caveats apply. We gratefully acknowledge the Italian Ministry of Education, University and Research (2008 Prin project on "Social Capital, Corporate Social Responsibility and Performance" – Padova unit) and the University of Verona (2011 Joint Projects on "Punishment and Decision-making: Neuroeconomic Foundations, Behavioural Experiments and Implications for Law and Economics") for financial support.

Quis custodiet ipsos custodes?

Juvenal

1. Introduction

A defining feature of modern democracies is the presence of third-party institutions serving the public interest by sanctioning wrongdoing. Judges are charged with the task of combating illegal behavior such as crime, drug trafficking and terrorism, and politicians are expected to enact laws that deter misconduct. Likewise, regulators are duty-bound to promote competition among firms and to take tough action against those acting non-competitively. In each of these cases, ordinary citizens bear the brunt of the costs associated with dishonest behavior and thus ultimately benefit most from the appropriate top-down measures taken by public officials against wrongdoers.

However, while it is true that a number of influential public posts are filled by appointment, this does not mean that citizens must always stand by passively and wait for delegated third parties to curb malfeasance. In contemporary democratic systems, citizens have at their disposal several means of exerting bottom-up pressure on third-party institutions and holding them *accountable*. In the U.S., most state judges are elected by the people. This creates a direct link between judges and the general populous that is lacking in states or countries where judges are appointed[1]. As for legislators, bottom-up pressure from the citizens can take many other forms than regular elections[2], from advocacy and lobbying activities on the part of organized groups of citizens to direct participation in reform processes (e.g., promotion of direct democracy initiatives such as referenda over issues deemed important). This is also increasingly true in developing countries, as the 2004 World Development Report claims that development interventions would be more effective if poor people were put "at the centre of service provision, by enabling them to monitor and discipline service providers, by amplifying their voice in policy making" (cited in Serra, 2012). The selection of regulators provides another interesting example. As Besley and Coate (2003) point out, while the heads of regulatory agencies are normally appointed by politicians, "a number of U.S. states have

injected a degree of populism into the regulatory process by requiring that the heads of their independent regulatory commissions be directly elected" (p. 1176). Hence, in real life we can easily find several instances of both accountable and nonaccountable public officials within various domains, with the former being subject to different forms of bottom-up pressure. Modern advances in technology have also provided citizens with new tools for influencing public officials. For example, the spread of information technology has sparked a 'new politics of the Internet', with growing political protest fueled by social media. As the Economist recently put it, groups such as the Tea Party, the Occupy movement and many others worldwide seek the fast-multiplying effect that the Internet can add to activism and uprisings[3]. Fact-checking websites are also an example of a modern form of bottom-up accountability: their purpose is to check the factual accuracy of statements by major political players, both during and after electoral campaigns, in the form of TV ads, speeches and interviews[4].

It is therefore plausible to conjecture that differences in the selection, retention or monitoring of public officials can lead to differences in their subsequent behavior. A key implication is that whether and how accountable and nonaccountable officials differ in their actions and performance is ultimately an empirical question. In this study, we have recourse to experimental economics to provide, to the best of our knowledge, the first rigorous test of the impact of bottom-up pressure on third-party institutions in a controlled laboratory setting[5]. Our workhorse is the Third-Party Punishment Game originally proposed by Fehr and Fischbacher (2004). While recourse to this experimental setting within the growing literature on the topic has been confined mainly to analysis of prosociality and reactions to violation of social norms (see Section 2.2 for a brief survey), here we opt for a different interpretation of this game protocol and use it to investigate the behavior of public officials (e.g., judges, regulators, or politicians) in their interactions with other players such as wrongdoers and ordinary citizens. The novel feature of our experimental design is that it allows us to manipulate the institutional variable and directly compare a 'non-accountable institutions scenario' – where third parties are unaffected by citizens' behavior

and their sanctioning decision can be viewed as a purely top-down intervention – with an 'accountable institutions scenario' – where citizens make their voice heard and exert bottom-up pressure on public officials. The main purpose of this article is to shed light on the sensitivity of third-party intervention to bottom-up monitoring by ordinary citizens, who are the major victims of wrongdoers' behavior. Our results indicate that when citizens make their voices heard, officials impose significantly higher sanctions for wrongdoing. This finding suggests that, other things equal, the efforts made by judges, regulators and politicians to fight malfeasance are greater when they are held accountable to their citizens.

The structure of the remainder of the paper is as follows. Section 2 presents the related literature. Section 3 illustrates our experimental design and procedure. In Section 4 our experimental hypotheses are derived. Section 5 contains our major findings. Section 6 discusses and concludes.

2. Related literature

Related papers include both fieldwork and lab experiments. Here we first review a selection of contributions based on field data focusing on the distinction between accountable vs. nonaccountable public officials (Section 2.1). Then, we highlight the most significant studies on third-party punishment conducted by experimental economists in recent years (Section 2.2).

2.1. Accountable vs. nonaccountable officials' behavior: evidence from the field

Numerous empirical studies conclude that accountable and nonaccountable officials do in fact behave differently. In studying judges' decision-making, Choi et al. (2010) find that although appointed judges write higher quality opinions than elected ones, the latter display higher productivity. Moreover, elected judges are less likely to favor criminal defendants (Pinello, 1995), more consistent in their opinions (Hanssen, 1999) and more likely to rule in favor of employees in

employment discrimination cases (Besley and Payne, 2013). Gordon and Huber's (2007) analysis of the sentencing behavior of district court judges in Kansas shows that judges in partisan competitive systems hand down harsher sentences than those in retention systems[6]. Elected justices are also more inclined to overturn decisions of the lower court and less likely to change their preconceived opinions about a case (Iaryczower et al., 2013).

As to regulatory policy, Besley and Coate (2003) contrast direct election with the political appointment of regulators and, by focusing on electricity prices, show that elected regulators enact more consumer-orientated policies[7]. In their study of politicians, Besley and Case (2003) present evidence that the decisions made by U.S. governors subject to term limits are different from those who are not. Some empirical papers reveal that voters are willing to punish political corruption or other misbehavior by politicians (see, e.g., Peters and Welch, 1980), especially in electoral districts where civic attitudes are shared and widespread (Nannicini et al., 2013) and when the media make corruption a socially salient issue (Chang et al., 2010).

However, while the rich, fieldwork-based literature provides convergent evidence that accountable and nonaccountable officials in various domains produce different public policy outcomes, a major challenge for these studies is that a huge number of confounding factors make it difficult to clearly identify the actual impact of bottom-up pressure on third parties' behavior. For instance, wide variation in state-level characteristics in the U.S. may be the main driving force behind differences in justices' behavior rather than their selection mechanism[8]. Instead, the experimental method has the advantage of lending itself to specific analysis of the relationship between third-party sanctioning and citizens' punishment of third parties in a controlled environment, by ruling out confounding variables that inevitably exist outside of the laboratory[9]. Another, related drawback of relying on field data is that individual studies may not be representative, since they are often based on single countries or states with idiosyncratic features, making it impossible to draw generalizations.

2.2. Third-party punishment: evidence from the lab

In the experimental laboratory, subjects have often proved to be willing to display altruistic or nonstrategic punishment, which has been defined as costly sanctioning not driven by (more or less sophisticated forms of) material payoff maximization. However, behavioral economics, so far, has mainly dealt with second-party altruistic punishment, by focusing on the 'vengeful' behavior of experimental subjects who had been directly hurt by other players (see e.g. the seminal work by Fehr and Gächter, 2000). In recent years, a growing number of scholars have been arguing for the importance of costly sanctioning activities not only on the part of second parties, but also on the part of 'uninvolved' third parties. Today the economics of third-party punishment – unlike second-party punishment studies – is still in its infancy, especially at the experimental level[10]. Notable exceptions are the works by Fehr and Fischbacher (2004), Takahashi et al. (2005), Ottone (2005; 2008), Bernhard et al. (2006), Henrich et al. (2006), Marlowe et al. (2008), Charness et al. (2008), Carpenter and Matthews (2009; 2010) and Lewisch et al. (2011).

The experimental studies of Bernhard et al. (2006), Henrich et al. (2006) and Marlowe et al. (2008) investigate the influence of cultural factors on third-party sanctioning. Charness et al. (2008) explore the effects of third-party intervention in different treatments of an Investment Game and find a strong and significant effect of this sanctioning mechanism. Ottone (2008) implements a design where the third party has the opportunity to both punish the Dictator and transfer money to the Receiver (Solomon's Game)[11] and finds that the third party's transfers to the Receiver appear to be complementary to the punishment of the Dictator at high levels of unfairness and to be substitutes of it at low levels. Lewisch et al. (2011) compare punishment by a single third party to punishment by two third parties in different scenarios where the cost of punishment changes. They show that third-party punishment is an ordinary good and that punishers are heterogeneous.

The aforementioned experimental studies provide useful contributions to our understanding of third-party punishment mainly as a means of enforcing social norms. Here we propose that experimental economics is also an ideal method of inquiry into the crucial question of how bottom-

up accountability impacts on third-party institutions' decision-making. If, as here, third parties are considered to be public officials acting in the public interest, it is natural to ask: do public officials alter their propensity to sanction misconduct when held accountable to their citizens? We answer this question in the following sections by implementing a novel experimental design that allows us to directly compare the sanctioning behavior of accountable vs. nonaccountable third parties.

3. Experimental setup

3.1. Experimental design

We provide a laboratory simulation of the following real-life situations:

1) a subject can harm someone who cannot react and neither social nor legal monitoring exists;
2) the victim(s) of an unfair action cannot react, but a third party has the power to sanction the wrongdoer;
3) the victim cannot directly punish the wrongdoer, but has the power to sanction the third party who could intervene against the wrongdoer;
4) a group of victims cannot directly punish the wrongdoer, but can decide by vote to sanction the third party who could intervene against the wrongdoer.

Our experiment aims to capture the real-life distinction between accountable and nonaccountable holders of public office, as outlined in Sections 1 and 2.1. Technically, the four situations above are captured by an experimental design that consists of five treatments (Figure 1): the Dictator Game Treatment (DG), the Third-party Punishment Game Treatment (TP), the Collective Third-party Punishment Game Treatment (CTP), the Accountable Third-party Punishment Game Treatment (ATP) and the Accountable Collective Third-party Punishment Game Treatment (ACTP)[12].

- FIGURE 1 HERE -

The Dictator Game Treatment (DG). Our tool here is the classic Dictator Game (hereafter, DG; see Forsythe et al., 1994). At the beginning of the session, each subject is randomly assigned a role (A or B) and groups of two participants are formed. Participants A receive 20 tokens and participants B 10 tokens. In each group, participant A and participant B play a DG. Player A can decide whether and how much money to transfer to player B by choosing a discrete number of tokens between 0 and 5[13]. Player B cannot react to any decision taken by player A. Hence, player A acts as the 'Dictator' and player B as the 'Receiver'.

The Third-party Punishment Game Treatment (TP). In this treatment, our vehicle is the 'third-party punishment in the dictator game' (hereafter, TP) originally proposed by Fehr and Fischbacher (2004), where a third player is introduced into the DG and given a punishment option. At the beginning of the first stage, each subject is randomly assigned a role (A, B or C) and groups of three participants are formed. In each group, participant A and participant B play a Dictator Game, acting as Dictator and Receiver, respectively. In the second stage, participant C (the Third Party) enters the game – endowed with 20 tokens like player A – and has to decide whether to bear a cost in order to sanction A or keep the whole initial endowment (with no costs). The cost for participant C to punish participant A by the amount of 2 tokens is 1 token. Only transfers of entire tokens are allowed and no participant can earn a negative payoff. In this setting, participant C is a potential third-party punisher: this design reflects democratic systems, where the power to sanction wrongdoing is typically delegated to third-party institutions.

The Accountable Third-party Punishment Game Treatment (ATP). Here, we introduce a novel variant of the TP, which introduces a third stage after players make their decisions about third-party punishment. In Stage 3, participant B can decide to become an active player by punishing participant C. The cost for participant B to punish participant C by the amount of 2 tokens, is 1 token. Again, only transfers of entire tokens are allowed and no participant can earn a negative

payoff. This design feature reflects the idea that in real life citizens often have the opportunity to speak out, so that third-party sanctioning institutions are held accountable (e.g., when judges or regulators are elected). As far as the existing experimental literature on punishment is concerned, the ATP may be viewed as a combination of the TP and the well-known Ultimatum Game (UG; see Güth et al., 1982), where, as in the TP, participant C can punish participant A at a cost and, as in the UG, participant B can punish participant A. The key difference between the ATP and the UG is that in the latter the Receiver can *directly* (though implicitly, that is by rejecting the offer) punish his co-player[14], whereas in the former the Receiver is only allowed to *indirectly* punish the first party by punishing the third party for not (sufficiently) punishing the first party.

The Collective Third-party Punishment Game Treatment (CTP). Here, we introduce another variant of the TP. Each group is made up of five participants, with one participant A, three participants B and one participant C[15]. The initial endowment is still 20 tokens for A and C and 10 tokens for each B. As in the first stage of the TP, participant A and participants B play a DG. For each token transferred by A, each B receives a token. As an example, if A transfers 2 tokens, each B receives 2 tokens. Thus, the final payment is 18 for A and 12 for each B. The second stage is exactly the same as in the TP. This treatment allows us to elicit how the Third Party reacts when an unfair action affects more than one subject. It also provides a useful benchmark for conducting rigorous analysis once we introduce the voting procedure.

The Accountable Collective Third-party Punishment Game Treatment (ACTP). This treatment combines some of the characteristics of the CTP with some of the ATP. As in the CTP, each group is made up of five subjects (one A, three Bs and one C). The first and the second stage are run as in the CTP, so each token transferred by A is received by each of the subjects B[16]. As in the ATP, participants B can sanction participant C during a third stage, so participants B end up voting twice. During the first ballot, they have to decide whether subject C should be sanctioned or not. If all

three Bs choose to sanction C, a second ballot takes place to determine the level of punishment, for which they have ten opportunities to make a unanimous decision. The cost for each participant B to punish participant C by the amount of 2 tokens, is 1 token. Participants B are aware that if they cannot reach a unanimous decision on the level of sanctioning, C will receive no sanction at all. The reason for insisting on unanimity in this treatment is that we wanted the three Receivers to focus on the (double) coordination problem (i.e., the decisions about whether and how much to sanction the Third Party), rather than on the classic free-rider problem. In other words, our experimental design rules out the possibility that one of the Bs might be tempted to leave the burden of sanctioning C to the other two. So either all participants B agree on whether and how much to punish C (and, therefore, share the associated costs equally) or C avoids being punished altogether.

3.2. Fairness reference point elicitation

Today the growing consensus among law and political science scholars is that judicial decision-making, far from being 'objective' and 'neutral', is influenced by the set of values subjectively held by judges[17]. In order to capture this inevitable subjectivity in justices and other third parties' decision-making, in this work we refer to Konow's (2001) theory of distributive justice. According to Konow, a disinterested observer's disutility can emerge from the difference between the actual allocation of resources and an ideal allocation that depends on a set of principles that are differently valued by the subjects. A key point in his theory is that different observers can have different reference points. This also implies inter alia that egalitarian allocation is only one of many different possible solutions. As argued by Konow, "equality is not a principle of fairness; at best it is a special case when members are equally accountable, efficient or needy" (p. 159).

In light of these arguments, we claim that since third parties' reaction to unfairness is likely to strictly depend on what is actually considered as fair by the players themselves, it would not be methodologically sound to take for granted, ex ante, that all the players in in a DG laboratory subjectively perceive as 'fair' the classic 50/50 norm – that is that they all share the so called

'egalitarian distribution norm'[18]. For this reason, after each subject makes a decision in each treatment during the experiment, we elicit information concerning the number of tokens the player thinks the Dictator ought to transfer to the Receiver (i.e., the players' normative beliefs), that is, the 'fair transfer'[19]. We define this as the 'Players' fairness reference point'. If we consider this ideal transfer as a subjective reference point of fairness, actual transfers by Dictators that are lower than the ideal transfer may be considered unfair by of the Third Party. If subjects C think that an equal share is the fair decision, the analysis will be the same as in Fehr and Fischbacher (2004): we will detect Third Parties' punishment choices as the gap between the transfer from the Dictator to the Receiver and the equal share increases. If this is not the case, it seems natural to analyze the Third Party's reaction when her subjective principle of fairness is violated[20].

Our experiment was therefore designed to allow us to detect whether subjects' fairness reference point is: 1) sensitive to the presence of a Third Party who can sanction the Dictator[21] or 2) common to all the players belonging to our artificial micro-society. By asking all players to state their fairness reference points in four different situations with increasing levels of control, we can suss out whether the Third Parties' notion of fairness is shared by the other subjects in different situations. Though our major focus is Third Parties' sense of fairness (and its relationship with their sanctioning behavior in different treatments), we view this as a relevant point: if institutions' ideas of fairness were strongly different from those of citizens, it would be a potential source of both dissatisfaction and conflict.

3.3. Experimental procedure

The experiment was run at the University of Milano-Bicocca (EELAB) and programmed with the software z-Tree (Fischbacher, 2007). Overall, 377 subjects participated in the experiment (40 participants in the DG, 60 in the TP, 57 in the ATP, 120 in the CTP and 100 in the ACTP – see Table 1), recruited through a web-based recruitment system. Each subject participated in one treatment only (between subjects design) and was only assigned just one role (Dictator, Receiver or

Third Party) within that treatment. Participants were not informed about the other treatments, and their partners' identities remained unknown even when the experiment was over. In all treatments, the games were one-shot: each player only played once, regardless of her role.

- TABLE 1 HERE -

At the beginning of the experiment, participants were informed about the sequential nature of the game protocol. The instructions were read by participants on their computer screen while an experimenter read them out loud. After reading the instructions and before the subjects were invited to make their decisions, some control questions (see Figure A.6 and Figure A.7) were asked in order to make sure that players understood the rules of the game. At the end of each session, we elicited the subjects' perception of fairness[22] and then the subjects were asked to fill out a brief questionnaire to check for socio-demographic data. In order to minimize framing effects, we presented the experiment in a context-free fashion. As it is customary in economic experiments, the language was neutral. For example, the instructions did not contain loaded verbs such as *punish* or *sanction*, but used instead the more neutral term *reduction;* players were labeled 'Participant A', 'Participant B' and 'Participant C', with no reference to wrongdoers, citizens or public officials[23].

The strategy method was implemented during the Third Party's stage, in line with previous work on third-party punishment (see Fehr and Fischbacher, 2004; Henrich et al., 2006; Almenberg et al., 2011)[24]. When the strategy method is used, subjects are asked to state their decision corresponding to each possible case. In our experiment, this meant that the Third Party was asked to indicate the number of points to subtract for each of the Dictator's possible transfer levels before knowing the Dictator's actual choice. The final payoff was then determined on the basis of the Dictator's actual choice. In order to help Third Parties think carefully about their decisions, an overview of the resulting payoffs was made available (see Figure A.5). Each session lasted for about 20 minutes for the DG, 40 minutes for the TP and the CTP, 50 minutes for the ATP and 60

minutes for the ACTP. Each subject earned on average 7 Euros, considering that each token was worth 50 eurocents.

4. Experimental hypotheses

Our design allows us to test two hypotheses concerning the Third Parties' behavior. The first deals with the Third Parties' reaction to unfairness. Standard economic theory predicts that, in a one-shot context, rational and self-interested Third Parties will never punish, whether in the TP or in the CTP: punishment is costly for them and brings no monetary benefits. For the same reason, the same behavior is also expected of both the Receiver and the Third Party in the ATP and in the ACTP[25]. By contrast, recently developed models on 'social preferences' suppose that people's behavior, far from being driven by material self-interest alone, is also influenced by motivations such as altruism (Levine, 1998), envy, inequity aversion (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) or reciprocity (Rabin, 1993; Falk et al., 2008). However, we chose to disregard this consideration in designing our experiment, since we believe that this type of motivation, which several laboratory studies have documented to be widespread in many peer-to-peer interactions, should not drive third-party institutions' decisional processes. By contrast, as already discussed in Section 3.2, we are interested in investigating whether punishment emerges as a reaction by Third Parties to differences between the actual level of transfer from the Dictator to the Receiver and their own reference point of fairness. This would be in line with Konow's (2001) theory of justice based on inequity aversion, according to which "an observer's utility is assumed to be a function of the difference, say, for member *i* between *i*'s actual allocation, denoted y_i, and his or her entitlement, denoted η_i. […] let this be represented by an inequity aversion term, $-f(y_i - \eta_i)$" (p. 141). In our experiment y_i is represented by the actual transfer from the Dictator to the Receiver and η_i captures Third Parties' beliefs concerning the number of tokens the Dictator ought to transfer to the Receiver. The first hypothesis is also based on the recent experimental literature on third-party

punishment mentioned in section 2.2: according to these works, Third Parties react to unfairness, and their level of punishment increases as the level of unfairness increases. Therefore:

Hypothesis 1 (H1): in all treatments where Third Parties are allowed to punish, their expenditure on punishment increases as the level of subjective unfairness increases.

The second hypothesis is based on the empirical findings on accountable officials' behavior reported in section 2.1. According to these studies, accountable officials in various domains seem more likely to act on the behalf of their citizens by sanctioning wrongdoing than nonaccountable officials. Therefore:

Hypothesis 2 (H2): in the ATP (resp., in the ACTP), Third Parties' expenditure on punishment is higher than in the TP (resp., in the CTP), for each level of subjective unfairness[26].

5. Experimental evidence

In this section we present our experimental evidence concerning the key issue of bottom-up accountability, by investigating the effect of citizens' pressure on third parties' choices. Our main findings can be summarized as follows:

Result: In all treatments where punishment is possible, Third Parties' levels of punishment are sensitive to their subjective sense of fairness. Moreover, when violations of subjective fairness occur, third-party punishment is harsher in the presence of bottom-up accountability.

Our analysis consisted of two steps:

a) the definition of the fairness reference point for Third Parties;

b) the comparison of the level of punishment chosen by Third Parties for each level of unfairness in all treatments where punishment is possible.

5.1. Definition of Third Parties' fairness reference point

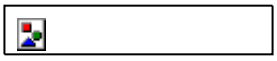
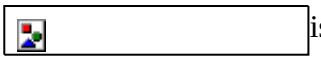
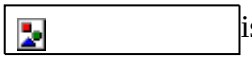
At the end of each session, we elicit Third Parties' beliefs concerning the number of tokens the Dictator *ought to* transfer to the Receiver (i.e. Third Parties' *normative* beliefs) and define it as the 'Third Parties' fairness reference point'. We find that most Third Parties expect a Dictator to give *something*. However, 10% of them at most think that the Dictator ought to give 1/4 of her endowment to the Receiver (see Table 2)[27]. A series of t-tests[28] is performed to check whether Third Parties consider the fair transfer from the Dictator to the Receiver as different from 0 and from 5. In both cases, p = 0.000[29]. Moreover, if we compare the average level of fair transfers according to the Third Parties, we find out that there is no significant difference along the treatments (Kruskall-Wallis test, $p = 0.17$)[30]. This means that Third Parties' ideal fair transfer is not influenced by the specific experimental game they play. In other words, it turns out that Third Parties believe that, if you happen to be playing as a Dictator, playing entirely selfishly is unfair. At the same time, they do not believe that splitting the pie equally is morally compulsory[31]. This result is in line with Konow's (2001) view that equality is only one specific version of fairness.

- TABLE 2 AND TABLE 3 HERE -

In light of this result, we will analyze Third Party's reaction when her subjective principle of fairness is violated.

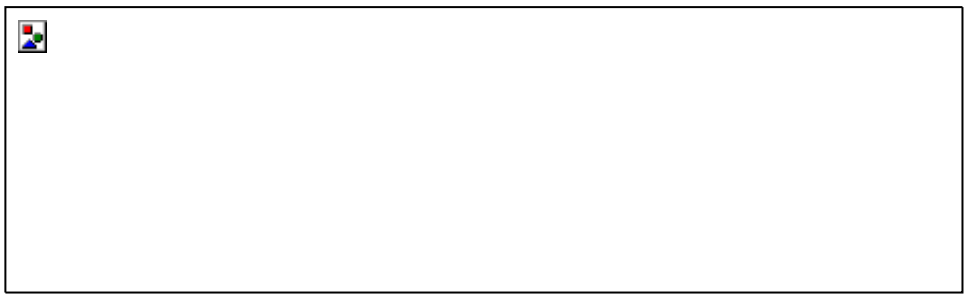
5.2. Third Parties' reaction to subjective unfairness and bottom-up accountability

We will consider two measures of subjective unfairness that register the distance between the reference point and the Dictator's transfer: the negative subjective unfairness, when the Dictator transfers less than the amount considered fair by the Third Party (ΔNEG = max {0, Fair Transfer – Dictator's Transfer}), and the positive subjective unfairness, which represents the converse situation (ΔPOS = max {0, Dictator's Transfer – Fair Transfer}). Since we implemented the strategy method at the third-party punishment stage, we can observe each Third Party's reaction to different levels of subjective unfairness[32]. Consequently, our dataset is a panel. This implies that Third Parties' reaction to subjective unfairness – which is measured by the number of tokens spent to punish the Dictator (P_i) – can be detected across the treatments through random-effects Tobit regressions censored at the low level 0 whose general specification is:

where:  represents a list of dummies equal to 1 if the observation belongs to the specific treatment *k* and 0 otherwise;  is a list of interactions between the TREATMENT *k* and ΔNEG_{ij} and ΔPOS_{ij} respectively;  is a list of demographic variables. More specifically, we ran three regressions – R1, R2 and R3. In R1 we analyze Third Parties' behavior in the 3-player scenarios (TP and ATP), while in R2 we perform the same analysis for the collective 5-player scenarios where each couple made up of a Dictator and a Third Party is assigned a group of three Receivers (CTP and ACTP). Finally, R3 has been run on the pooled sample. Consequently:

(R1)

(R2)



(R3)

where:

CTP_i is a dummy variable equal to 1 if the observation belongs to the CTP;

ATP_i is a dummy variable equal to 1 if the observation belongs to the ATP;

$ACTP_i$ is a dummy variable equal to 1 if the observation belongs to the ACTP;

$\Delta NEG*CTP_{ij}$, $\Delta NEG*ATP_{ij}$, $\Delta NEG*ACTP_{ij}$, are variables equal to max {0, Fair Transfer – Dictator's Transfer} if the observation belongs to the CTP, the ATP and to the ACTP respectively;

$\Delta POS*CTP_{ij}$, $\Delta POS*ATP_{ij}$, $\Delta POS*ACTP_{ij}$, are variables equal to max {0, Dictator's Transfer – Fair Transfer} if the observation belongs to the CTP, the ATP and to the ACTP respectively;

$MALE_i$ is a dummy variable equal to 1 if the subject is a male;

$FIRST_i$ is a dummy variable equal to 1 if the subject is a first-time participant in an experiment;

JOB_i is a dummy variable equal to 1 if the subject has a job.

Evidence 1: Third Parties' reaction increases as Dictators' unfairness increases.

In Figure 2 and in Table 4 the relation between the (subjectively perceived) unfairness of Dictators and the level of punishment from the Third Party to the Dictator in each treatment is depicted.

- FIGURE 2 AND TABLE 4 HERE -

It is clear that the level of punishment increases as the Dictator's transfer becomes increasingly lower than that considered fair by the Third Party: Third Parties' perceived sense of fairness is *reference-dependent*. All regressions confirm the existence of a positive relation between the level of punishment and the degree of 'negative subjective unfairness' (ΔNEG_{ij}, $p = 0.000$ both in R1 and R3; $p = 0.002$ in R2). From Figure 2 it also emerges that some punishment still exists even when the Dictator transfers to the Receiver a sum that is higher than the Third Party's fair transfer. Nevertheless, our econometric analysis indicates that the quantitative relevance of this phenomenon is not significant. Moreover, the constant term in all regressions is not significantly different from zero. This means that when the Dictator transfers what is considered to be a fair amount of tokens, Third Parties do not intervene. Finally, the coefficient related to the variable 'positive subjective unfairness' (ΔPOS_{ij}) is negative. This suggests that when the Dictators' transfer is higher than the fairness reference point, Third Parties are ready to give them a reward for this. See Table 5 for details.

Observation 1. Our findings provide support for Hypothesis 1. In all treatments where punishment on the part of Third Parties is possible, Third Parties react to unfairness, even though they are not the victims of Dictators' unfair behavior. This is consistent with the experimental literature on third-party punishment summarized in Section 2.2.

- TABLE 5 HERE -

Evidence 2: Third Parties' reaction to subjective unfairness is stronger in the ATP (resp., in the ACTP) than in the TP (resp., in the CTP).

When we introduce the ATP treatment, departing from the classic Third-Party Punishment protocol and adding the possibility for the Receiver to punish the Third Party (e.g., if the Receiver thinks the Third Party did not sanction an unfair Dictator enough), it emerges that, as the Dictator's transfer becomes increasingly lower than that considered fair by the Third Party, the Third Party's reaction is significantly stronger than in the TP (ΔNEG*ATP, p = 0.000 in R1; see Figure 2). Likewise, we find that Third Parties' reaction to subjective unfairness is stronger in the ACTP than in the CTP (ΔNEG*ACTP, p = 0.013 in R2): even in the presence of a group of three Receivers (rather than of a single Receiver), the possibility for Receivers to hold Third Parties accountable turns out to make a difference and induces Third Parties to react more strongly to subjective unfairness than in treatments where there is no bottom-up accountability.

Hence, accountable Third Parties punish more severely than nonaccountable ones. The percentage of punishers among the treatments is not significantly different (chi2 tests, p>0.10).

Observation 2. Our findings provide support for Hypothesis 2. Accountable Third Parties are more apt to punish wrongdoers than nonaccountable ones. This occurs under both individual (a single Receiver present) and group decision-making (three Receivers present). Moreover, while this difference occurs at the behavioral level, we find no significant differences in terms of the fairness reference point across treatments. This suggests that pressure from citizens does *not* alter Third Parties' notion of *fairness*, whereas it does have an impact on their *behavioral* reaction to unfairness.

5.3. Ancillary Result

Evidence 3: In the ACTP, where a group of Receivers may punish Third Parties, Third Parties punish negative subjective unfairness less severely than in the ATP, where a single Receiver is present.

In the ATP, a single Receiver holds the Third Party accountable, whereas in the ACTP a group of three players acting as Receivers need to unanimously agree upon whether and how much to sanction the Third Party. As explained in Section 3.1, we opted for unanimity rule in the ACTP, without which the Third Party would not be punished so as to exclude the free-rider problem. We believe that the need for the players to reach unanimity in the two ballots makes our rule a rather demanding coordination device. Hence, it is plausible to believe that Third Parties will anticipate this and act accordingly, by punishing significantly more severely in the ATP than in the ACTP. Similarly, we also speculatively argue that this difference in the punishment behavior of the Third Party across the two treatments would be mitigated in the presence of a less extreme coordination device in the ACTP.

6. Discussion and conclusions

The primary aim of this paper was to determine whether public officials responsible for top-down interventions to curb malfeasance are sensitive to bottom-up pressure on the part of ordinary citizens, the primary victims of wrongdoing. Our experimental investigation shows that when citizens are proactive and make their voices heard, third-party institutions impose more severe sanctions. This finding lends itself naturally to several observations.

Our main result is that accountable third parties are more willing to listen to citizens' voice than nonaccountable third parties. But is this good news or bad news for society as a whole? Agency models view the judiciary or individual judges as agents, and the public or particular elected officials as the principals (e.g., Maskin and Tirole, 2004). These models predict that, unless properly selected, monitored and rewarded, agents will not act in the interests of the principals. What is good about systems where judges are elected is that fear of sanctions can induce them to be more productive: the empirical works cited in Section 2.1, as well as our experiment, confirm that

this is actually the case. However, several serious potential drawbacks to this reaction to citizens' pressure have to be considered. Let us turn to some of the 'dark sides' of bottom-up accountability.

First, elected judges may rule on cases in partisan fashion in order to avoid the risk of a partisan sanction (see Choi et al., 2007; see on this also Garoupa et al., 2013). Politicians may also act in partisan fashion, at the same time running the risk of generating sizable groups of dissatisfied minorities. In this regard, Maskin and Tirole's (2004) model predicts that even though, on the one hand, "accountability allows the public to screen and discipline their officials; on the other, it may induce those officials to pander to public opinion and put too little weight on minority welfare" (p. 1034). In other words, officials may overlook minority interests and choose actions that are *popular*, rather than those that are right for society (see on this also Iaryczower et al., 2013). When the majority's preferences are very likely to inflict large negative externalities on the minority, nonaccountability is preferable to accountability (Maskin and Tirole, 2004). As Maskin and Tirole rightly point out, although some people might view such pandering as 'responsiveness', it is in clear conflict with the rationale for representative democracy, in which the delegation of power to representatives typically occurs because the latter are expected to do a better job than ordinary citizens: "As specialists in public decision-making, they are more likely than the average citizen to have the experience, judgment, and information to decide wisely" (Maskin and Tirole, 2004; p. 1034).

This naturally leads us to a second, related disadvantage to bottom-up accountability: citizens may lack the necessary information to make wide decisions about the actions third-party institutions should take. It has been argued that in some contexts citizens are liable to have better information than other players: for example, in developing countries characterized by high levels of widespread corruption, citizens may be more informed than others about the quality of services provided and about corruption, if they are being bribed (Serra, 2012). However, in many other cases it is hardly deniable that the electorate is poorly informed about the optimal actions to be taken (see on this Reinikka and Svensson, 2004)[33], so that, in such situations, as noted by Maskin and Tirole

(2004), nonaccountability is better than accountability. Hence, technical decisions may be best allocated to appointed rather than elected officials (see on this also Alesina and Tabellini, 2007). Iaryczower et al.'s (2013) empirical findings on criminal decisions across U.S. states' Supreme Courts confirm that appointed justices have better information on average and are more likely to change their preconceived opinions about a case (i.e. they have greater flexibility) and make fewer mistakes than their elected counterparts (i.e. they are less likely to make an incorrect ruling). It is clear that important trade-offs need to be considered and empirically addressed when considering between appointed and accountable public officials.

Thirdly, the capacity for citizens to hold public officials accountable depends crucially on the nature and level of social capital prevailing in the community under study. In areas where social capital is high and takes the form of 'civic capital' (see Guiso et al.'s (2011) definition), citizens are willing to discipline politicians in order to promote the common good of their community, rather than their own personal interests (on this topic, see Nannicini et al. (2013) referred to in Section 2.1). By contrast, it is reasonable to think that in communities with low levels of civic capital, special interest groups will compete to influence politicians in desired ways. The analysis of bottom-up accountability in the presence of multiple groups of citizens, together with the comparison of different scenarios emerging depending on the level of social capital in the community, are left as interesting avenues for future research. Another step to be taken in future work will be the investigation of bottom-up accountability within a *repeated game* protocol, to see whether our results extend to such a new setting, where both strategic and nonstrategic motives for punishment are likely to play an important role[34].

Appendix – Instructions

Instructions for the Dictator Game Treatment

1.A&B. Welcome to the experiment, and thank you for participating. Please follow the instructions that will appear on your screen. There is nothing complicated, and there are no trick questions. Your answers will be absolutely anonymous. It will not be possible for the experimenters to match the answers with the person who provided them. For the success of the experiment, you must not communicate with one another.

2.A&B. The experiment involves two different kinds of participants – Participant A and Participant B. At the beginning of the experiment you will be randomly assigned a role (A or B) and you will be randomly paired with another participant. You will never know the identity of your partner.
Participant A has an initial endowment of 20 tokens. She has to decide on the number of tokens – from 0 to 5 – she wants to transfer to Participant B. Participant B has an initial endowment of 10 tokens and she cannot make any decisions.
Each token corresponds to 50 eurocents.

3.A&B. Payoffs are computed in the following way:
- Participant A's payoff is the initial endowment of 20 tokens minus the number of tokens she transfers to Participant B;
- Participant B's payoff is the initial endowment of 10 tokens plus the number of tokens she receives from Participant A.

4.A&B. In particular, it should be clear that each token transferred by Participant A to Participant B increases Participant's payoff by 1 token – that is, by 50 eurocents.

5.A&B. Control questions – see Figure A.1 in the Appendix

6.A. Decision screen for Participant A – see Figure A.2
6.B. Waiting screen for Participant B – Participant A is making her choice. Please, wait.

7.A&B. In general, how many tokens do you think Participant A should transfer to Participant B?

8.A&B. Payoff screen

9.A&B. Please fill in this brief questionnaire. Afterwards, the experiment will come to an end and you will receive your payment.

Instructions for the Third-party Punishment Game Treatment

1.A&B&C. Welcome to the experiment, and thank you for participating. Please follow the instructions that will appear on your screen. There is nothing complicated, and there are no trick questions. Your answers will be absolutely anonymous. It will not be possible for the experimenters to match the answers with the person who provided them. For the success of the experiment, you must not communicate with one another.

2.A&B&C. The experiment involves three different kinds of participants – Participant A, Participant B and Participant C – and it has a two-stage structure. At the beginning of the first stage

you will be randomly assigned a role (A, B or C) and you will be randomly paired with two other participants. You will never know the identity of your partners.
During the FIRST STAGE, Participant A has an initial endowment of 20 tokens. She has to decide on the number of tokens – from 0 to 5 – she wants to transfer to Participant B. Participant B has an initial endowment of 10 tokens and she cannot make any decisions.
During the SECOND STAGE, Participant C has an initial endowment of 20 tokens. She can either reduce Participant A's payoff at a cost – she has to spend 1 token to reduce Participant A's payoff by 2 tokens – or keep the whole endowment.
To sum up, Participant C can allocate her endowment any way she wants by taking into account that the total number of tokens she spends to reduce Participant A's payoff plus those she keeps for herself cannot exceed 20 tokens.
Participant C is asked to declare her choice for each possible transfer from Participant A to Participant B. The final payoffs are computed on the basis of the actual transfer from Participant A to Participant B. Each token corresponds to 50 eurocents.

3.A&B&C. Check whether it is clear that:
- if Participant A transfers 2 tokens to Participant B and Participant C spends 2 tokens to reduce Participant A's payoff, the final payoff will be 14 tokens for Participant A, 12 tokens for Participant B and 18 tokens for Participant C;
- if Participant A transfers 2 tokens to Participant B and Participant C keeps her whole endowment, the final payoff will be 18 tokens for Participant A, 12 tokens for Participant B and 20 tokens for Participant C;

4.A&B&C. In particular, it should be clear that:
- each token transferred by Participant A to Participant B increases Participant B's payoff by 1 token – that is, by 50 eurocents;
- each token – that is, for each 50 eurocents – spent by Participant C reduces Participant A's payoff by 2 tokens – that is, by 1 euro.

5.A&B&C. Control questions – see Figure A.3

6.A. Decision screen for Participant A – see Figure A.4
6.B. Waiting screen for Participant B – Participant A is making her choice. Please wait.
6.C. Decision screen for Participant C – see Figure A.5

7.A&B&C. In general, how many tokens do you think Participant A should transfer to Participant B?

8.A&B&C. Payoff screen

9.A&B&C. Please fill in this brief questionnaire. Afterwards, the experiment will come to an end and you will receive your payment.

Instructions for the Accountable Third-party Punishment Game Treatment

1.A&B&C. Welcome to the experiment, and thank you for participating. Please follow the instructions that will appear on your screen. There is nothing complicated, and there are no trick questions. Your answers will be absolutely anonymous. It will not be possible for the experimenters to match the answers with the person who provided them. For the success of the experiment, it is necessary that you do not communicate with each other.

2.A&B&C. The experiment involves five people and three different kinds of participants – one Participant A, three Participants B and one Participant C – and it has a three-stage structure. At the beginning of the first stage you will be assigned a role randomly (A, B or C) and you will be randomly paired with four other participants. You will never know the identity of your partners.
During the FIRST STAGE, Participant A has an initial endowment of 20 tokens. She has to decide on the number of tokens – from 0 to 5 – she wants to transfer to each Participant B and she knows that each token she decides to transfer will turn into an additional token for EACH Participant B. Participant B has an initial endowment of 10 tokens and she cannot make any decisions at this stage.
During the SECOND STAGE, Participant C has an initial endowment of 20 tokens. She can either reduce Participant A's payoff at a cost – she has to spend 1 token to reduce Participant A's payoff by 2 tokens – or keep the whole endowment.
To sum up, Participant C can allocate her endowment any way she wants by taking into account that the total number of tokens she spends to reduce Participant A's payoff plus those she keeps for herself cannot exceed 20 tokens.
Participant C is asked to declare her choice for each possible transfer from Participant A to Participant B. The final payoffs are computed on the basis of the actual transfer from Participant A to Participant B.
During the THIRD STAGE, Participant B is endowed with her initial endowment of 10 tokens plus the amount of tokens transferred by Participant A. She can either reduce Participant C's payoff at a cost – she has to spend 1 token to reduce Participant C's payoff by 2 tokens – or keep the whole endowment.
Each token corresponds to 50 eurocents.

3.A&B&C. Check whether it is clear that:
- if Participant A transfers 2 tokens to Participant B, Participant C spends 2 tokens to reduce Participant A's payoff and Participant B spends 1 token to reduce Participant C's payoff, the final payoff will be 14 tokens for Participant A, 11 tokens for Participant B and 16 tokens for Participant C;
- if Participant A transfers 2 tokens to Participant B, Participant C keeps her whole endowment and Participant B does not spend any tokens to reduce Participant C's payoff, the final payoff will be 18 tokens for Participant A, 12 tokens for Participant B and 20 tokens for Participant C;

4.A&B&C. In particular, it should be clear that:
- each token transferred by Participant A to Participant B increases Participant B's payoff by 1 token – that is, by 50 eurocents;
- each token – that is, for each 50 eurocents – spent by Participant C reduces Participant A's payoff by 2 tokens – that is, by 1 euro;
- each token – that is, for each 50 eurocents – spent by Participant B reduces Participant C's payoff by 2 tokens – that is, by 1 euro.

5.A&B&C. Control questions – see Figure A.6 and Figure A.7

6.A. Decision screen for Participant A – see Figure A.4
6.B. Decision screen for Participant B – Your role is B. Participant A had an initial endowment of 20 tokens. She transferred XXX tokens to you. Participant C had an initial endowment of 20 tokens. She spent ZZZ tokens to reduce Participant A's payoff. You are endowed with your initial endowment of 10 tokens plus the amount of tokens transferred by Participant A. You can either reduce Participant C's payoff at a cost – you have to spend 1 token to reduce Participant C's payoff

by 2 tokens – or keep your whole endowment. How many tokens would you like to spend to reduce Participant C's payoff?
6.C. Decision screen for Participant C – see Figure A.5

7.A&B&C. In general, how many tokens do you think Participant A should transfer to Participant B?

8.A&B&C. Payoff screen

9.A&B&C. Please fill in this brief questionnaire. Afterwards, the experiment will come to an end and you will receive your payment.

Instructions for the Collective Third-party Punishment Game Treatment

1.A&B&C. Welcome to the experiment, and thank you for participating. Please follow the instructions that will appear on your screen. There is nothing complicated, and there are no questions. Your answers will be absolutely anonymous. It will not be possible for the experimenters to match the answers with the person who provided them. For the success of the experiment, it is necessary that you do not communicate with each other.

2.A&B&C. The experiment involves five people and three different kinds of participants – one Participant A, three Participant B and one Participant C – and it has a two-stage structure. At the beginning of the first stage you will be randomly assigned a role (A, B or C). You will never know the identity of your partners.
During the FIRST STAGE, Participant A has an initial endowment of 20 tokens. She has to decide on the number of tokens – from 0 to 5 – she wants to transfer to each Participant B and she knows that each token she decides to transfer will turn into an additional token for EACH Participant B. Each Participant B has an initial endowment of 10 tokens and she cannot make any decisions.
During the SECOND STAGE, Participant C has an initial endowment of 20 tokens. She can either reduce Participant A's payoff at a cost – she has to spend 1 token to reduce Participant A's payoff by 2 tokens – or keep the whole endowment.
To sum up, Participant C can allocate her endowment any way she wants by taking into account that the total number of tokens she spends to reduce Participant A's payoff plus those that she keeps for herself cannot exceed 20 tokens.
Participant C is asked to declare her choice for each possible transfer from Participant A to Participants B, knowing that A may transfer 0, 1, 2, 3, 4 or 5 tokens. The final payoffs are computed on the basis of the actual transfer from Participant A to Participants B, taking into account that each token corresponds to 50 eurocents.

3.A&B&C. Check whether it is clear that:
- if Participant A transfers 2 tokens to Participants B and Participant C spends 2 tokens to reduce Participant A's payoff, the final payoff will be 14 tokens for Participant A, 12 tokens for each Participant B and 18 tokens for Participant C;
- if Participant A transfers 2 tokens to Participants B and Participant C keeps her whole endowment, the final payoff will be 18 tokens for Participant A, 12 tokens for each Participant B and 20 tokens for Participant C;

4.A&B&C. In particular, it should be clear that:
- each token transferred by Participant A to Participants B increases the payoff of each Participant B by 1 token – that is, by 50 eurocents;

- each token – that is, for each 50 eurocents – spent by Participant C reduces Participant A's payoff by 2 tokens – that is, by 1 euro.

The subsequent instruction screens were similar to those used for the Third-party Punishment Game treatment described above.

Instructions for the Accountable Collective Third-party Punishment Game Treatment

1.A&B&C. Welcome to the experiment, and thank you for participating. Please follow the instructions that will appear on your screen. There is nothing complicated, and there are no questions. Your answers will be absolutely anonymous. It will not be possible for the experimenters to match the answers with the person who provided them. For the success of the experiment, it is necessary that you do not communicate with each other.

2.A&B&C. The experiment involves five people and three different kinds of participants – a Participant A, three Participants B and a Participant C – and it has a three-stage structure. At the beginning of the first stage you will be randomly assigned a role (A, B or C). You will never know the identity of your partners.
During the FIRST STAGE, Participant A has an initial endowment of 20 tokens. She has to decide on the number of tokens – from 0 to 5 – that she wants to transfer to each Participant B and she knows that each token she decides to transfer will turn into an additional token for EACH Participant B. Each Participant B has an initial endowment of 10 tokens and she cannot make any decisions at this stage.
During the SECOND STAGE, Participant C has an initial endowment of 20 tokens. She can either reduce Participant A's payoff at a cost – she has to spend 1 token to reduce Participant A's payoff by 2 tokens – or keep the whole endowment.
To sum up, Participant C can allocate her endowment any way she wants by taking into account that the total number of tokens she spends to reduce Participant A's payoff plus those she keeps for herself cannot exceed 20 tokens.
Participant C is asked to declare her choice for each possible transfer from Participant A to Participants B, knowing that A may transfer 0, 1, 2, 3, 4 or 5 tokens. The final payoffs are computed on the basis of the actual transfer from Participant A to Participants B, taking into account that each token corresponds to 50 eurocents.
During the THIRD STAGE, Participants B are endowed with their initial endowment of 10 tokens plus the number of tokens transferred to them by Participant A. Participants B can either reduce Participant C's payoff at a cost – they have to spend 1 token to reduce Participant C's payoff by 2 tokens – or keep the whole endowment.
In order to reduce Participant C's payoff, Participants B have to reach an agreement through a two-step voting procedure.

1) In the first ballot, Participants B have to decide on whether or not to reduce Participant C's payoff. If Participants B do not reach unanimity, Participant C's payoff remains unchanged and each Participant B keeps her initial endowment of 10 tokens plus the tokens transferred to her by Participant A in the first stage. In this case, the second ballot does not take place. If Participants B unanimously decide to reduce Participant C's payoff, then the second ballot takes place.
2) In the second ballot, Participants B who in the first ballot unanimously decided to reduce Participant C's payoff have to come to an agreement about how many tokens to spend to reduce Participant C's payoff. To do this, Participants B will have 10 attempts at their disposal and during each attempt each Participant B can suggest how many tokens Participants B should spend to reduce Participant C's payoff. Once they reach an agreement,

the chosen amount of tokens to be spent will be equally split among the three Participants B. By contrast, if Participants B do not reach an agreement after 10 attempts, the payoff of Participant C remains unchanged and each Participant B keeps their initial endowment of 10 tokens plus the tokens transferred to her by Participant A in the first stage.

3.A&B&C. Check whether it is clear that:
- if Participant A transfers 2 tokens to Participants B, Participant C spends 2 tokens to reduce Participant A's payoff and Participants B decide to spend a total of 3 tokens to reduce Participant C's payoff, the final payoff will be 14 tokens for Participant A, 11 tokens for each Participant B and 12 tokens for Participant C;
- if Participant A transfers 2 tokens to each Participant B, Participant C keeps her whole endowment and Participants B do not spend any tokens to reduce Participant C's payoff, the final payoff will be 18 tokens for Participant A, 12 tokens for each Participant B and 20 tokens for Participant C;

4.A&B&C. In particular, it should be clear that:
- each token transferred by Participant A to Participants B increases the payoff of each Participant B by 1 token – that is, by 50 eurocents;
- each token – that is, for each 50 eurocents – spent by Participant C reduces Participant A's payoff by 2 tokens – that is, by 1 euro;
- each token – that is, for each 50 eurocents – spent by Participants B reduces Participant C's payoff by 2 tokens – that is, by 1 euro.

The subsequent instruction screens were similar to the ones used for the Accountable Third-party Punishment Game treatment described above.

Figures and Tables

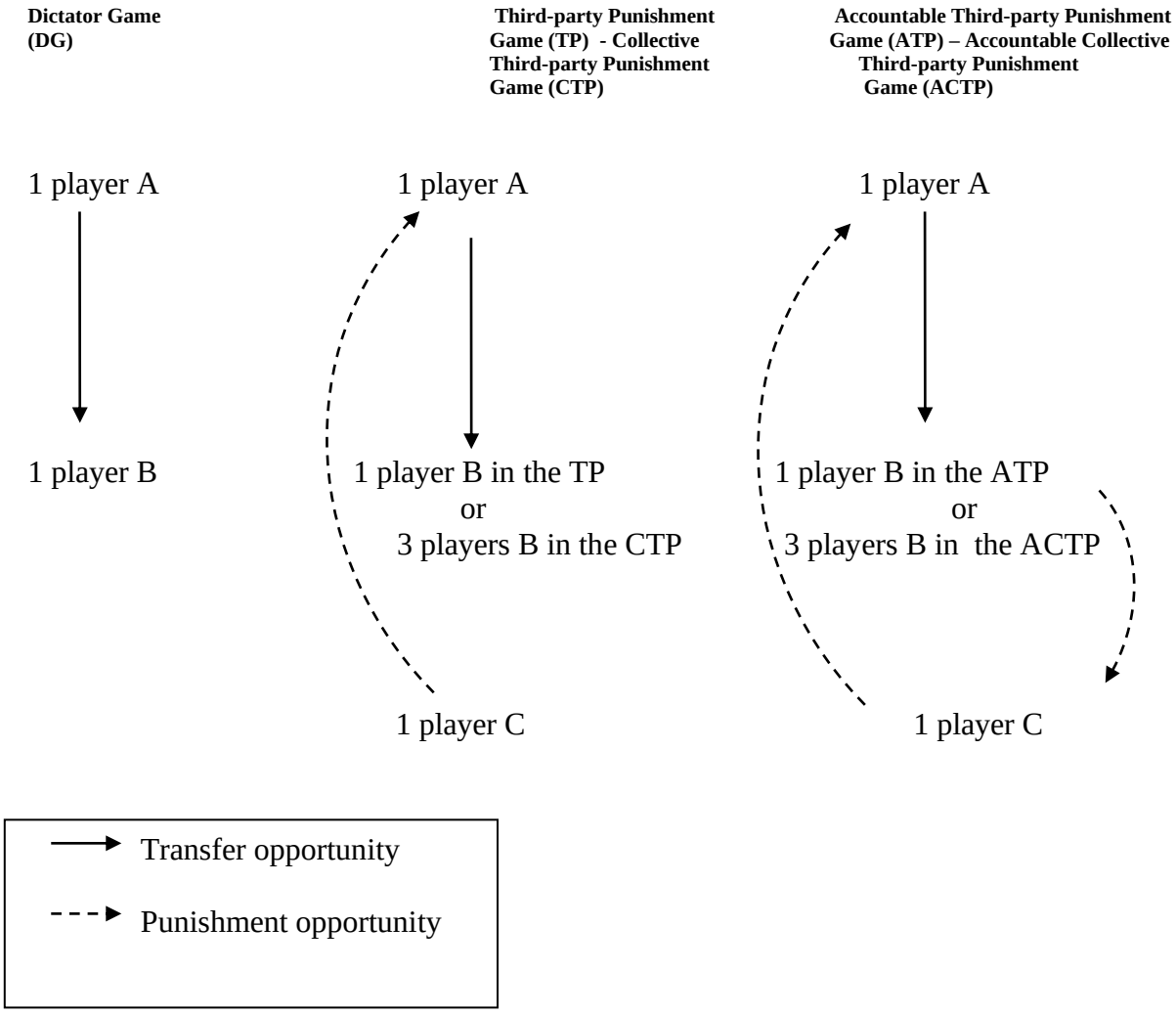


Fig. 1 The Five-Treatment Experimental Setup

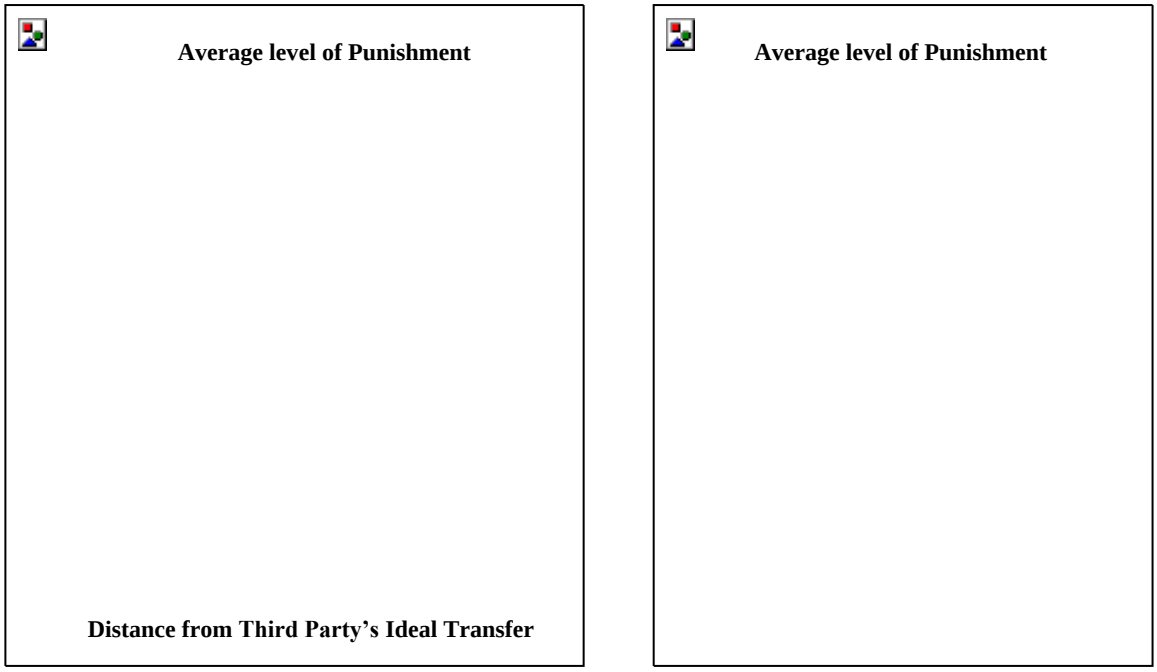


Fig. 2 Third Party's behavior

Table 1 Experimental sessions

	Third-party punishment	Accountability	Multiple Receivers	Observations
DG	NO	NO	NO	20 As, 20 Bs (20 groups)
TP	YES	NO	NO	20 As, 20 Bs, 20 Cs (20 groups)
ATP	YES	YES	NO	19 As, 19 Bs, 19 Cs (19 groups)
CTP	YES	NO	YES	24 As, 72 Bs, 24 Cs (24 groups)
ACTP	YES	YES	YES	20 As, 60 Bs, 20 Cs (20 groups)

Table 2. Distribution of actual and subjectively fair transfers across treatments

	Percentage of cases of	0	1	2	3	4	5
DG	Dictators who transfer…	40%	15%	25%	10%	5%	5%
	Dictators who think the fair transfer is …	30%	20%	30%	15%	0%	5%
	Receivers who think the fair transfer is…	5%	10%	30%	15%	5%	35%
	Observers who think the fair transfer is …	-	-	-	-	-	-
TP	Dictators who transfer…	30%	20%	25%	20%	0%	5%
	Dictators who think the fair transfer is …	30%	10%	25%	25%	0%	10%
	Receivers who think the fair transfer is …	10%	0%	40%	15%	0%	35%
	Observers who think the fair transfer is …	25%	5%	30%	30%	5%	5%
ATP	Dictators who transfer…	21%	42%	21%	0%	11%	5%
	Dictators who think the fair transfer is …	32%	26%	32%	0%	0%	11%
	Receivers who think the fair transfer is …	26%	11%	32%	11%	0%	21%
	Observers who think the fair transfer is …	47%	11%	16%	11%	11%	5%
CTP	Dictators who transfer…	38%	17%	17%	14%	10%	4%
	Dictators who think the fair transfer is …	35%	17%	24%	10%	10%	4%
	Receivers who think the fair transfer is …	29%	6%	25%	9%	6%	25%
	Observers who think the fair transfer is …	46%	4%	46%	0%	4%	0%
ACTP	Dictators who transfer…	44%	13%	30%	9%	0%	4%
	Dictators who think the fair transfer is …	48%	0%	26%	13%	0%	13%
	Receivers who think the fair transfer is …	23%	0%	22%	26%	9%	20%
	Observers who think the fair transfer is …	35%	0%	40%	15%	0%	10%

Table 3. Average actual and subjectively fair transfers across treatments

	The Dictator transfers to the Receiver (1)	**The Dictator thinks it is fair to transfer to the Receiver (2)**	**The Receiver thinks it is fair to transfer to the Receiver (3)**	**The Third Party thinks it is fair to transfer to the Receiver (4)**	***t- test (2) = (3) (3) = (4) (2) = (4)***
DG	1.4	1.5	3.1		
TP	1.55	1.85	3	2	
ATP	1.53	1.42	2.1	1.42	*p =0 .000*
CTP	1.52	1.55	2.33	1.12	*p = 0.000*
ACTP	1.22	1.56	2.58	1.75	*p =0 .919*
Kruskall-Wallis test DG = TP = ATP = CTP = ACTP	*p = 0.92*	*p = 0.88*	*p = 0.24*	*p = 0.17*	

Table 4. Average level of punishment expenditure for each level of subjective unfairness across treatments

	When the subjective unfairness is …										
Average Punishment is … (in tokens)	**-5**	**-4**	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**	**4**	**5**
TP	5	4.5	2.62	1.79	1.4	1.2	1.26	1.05	1.17	0.33	0
ATP	10	7.67	5	2.88	1.9	0.84	0.61	0.25	0.21	0.36	0.22
CTP	-	4	3	2	1.69	0.87	0.79	0.61	0.43	0.17	0.09
ACTP	4.5	4	4.2	2.69	1.69	1.05	0.72	0.67	0.47	0.71	0.71

Table 5. The econometric analysis. Dependent variable: tokens spent on punishment

	R1	**R2**	**R3**
ΔNEG	0.638***	0.553**	0.628***
	(0.175)	(0.178)	(0.150)
ΔPOS	-0.264*	-0.391***	-0.241*
	(0.157)	(0.111)	(0.134)
ATP	-1.83*	-	-1.515
	(1.084)		(0.922)
CTP	-	-	0.537
			(0.904)
ACTP	-	-0.471	0.111
		(0.698)	(0.896)
*ΔNEG*ATP*	1.047***	-	1.022***
	(0.268)		(0.230)
*ΔPOS*ATP*	0.082	-	0.093
	(0.247)		(0.212)
*ΔNEG*CTP*	-	-	-0.080
			(0.258)
*ΔPOS*CTP*	-	-	-0.178
			(0.187)
*ΔNEG*ACTP*	-	0.544**	0.501**
		(0.220)	(0.214)
*ΔPOS*ACTP*	-	0.162	-0.001
		(0.156)	(0.188)
AGE	0.314	-0.168	-0.040
	(0.287)	(0.149)	(0.149)
MALE	0.372	0.204	0.060
	(1.098)	(0.691)	(0.651)
FIRST	3.438**	2.171***	2.546***
	(1.356)	(0.665)	(0.686)
JOB	0.984	-0.252	0.435
	(1.058)	(0.846)	(0.692)
Constant	-9.508	2.692	-1.132
	(6.348)	(3.125)	(3.147)
N	39	44	83
T	6	6	6
N	234	264	498
Log Likelihood	253.2827	-254.7036	515.6903
Sigma_u	2.69***	1.74***	2.34***
Sigma_e	1.46***	1.08***	1.26***
Left-censored obs.	134	144	278

****significance 1%* ***significance 5%* **significance 10%*

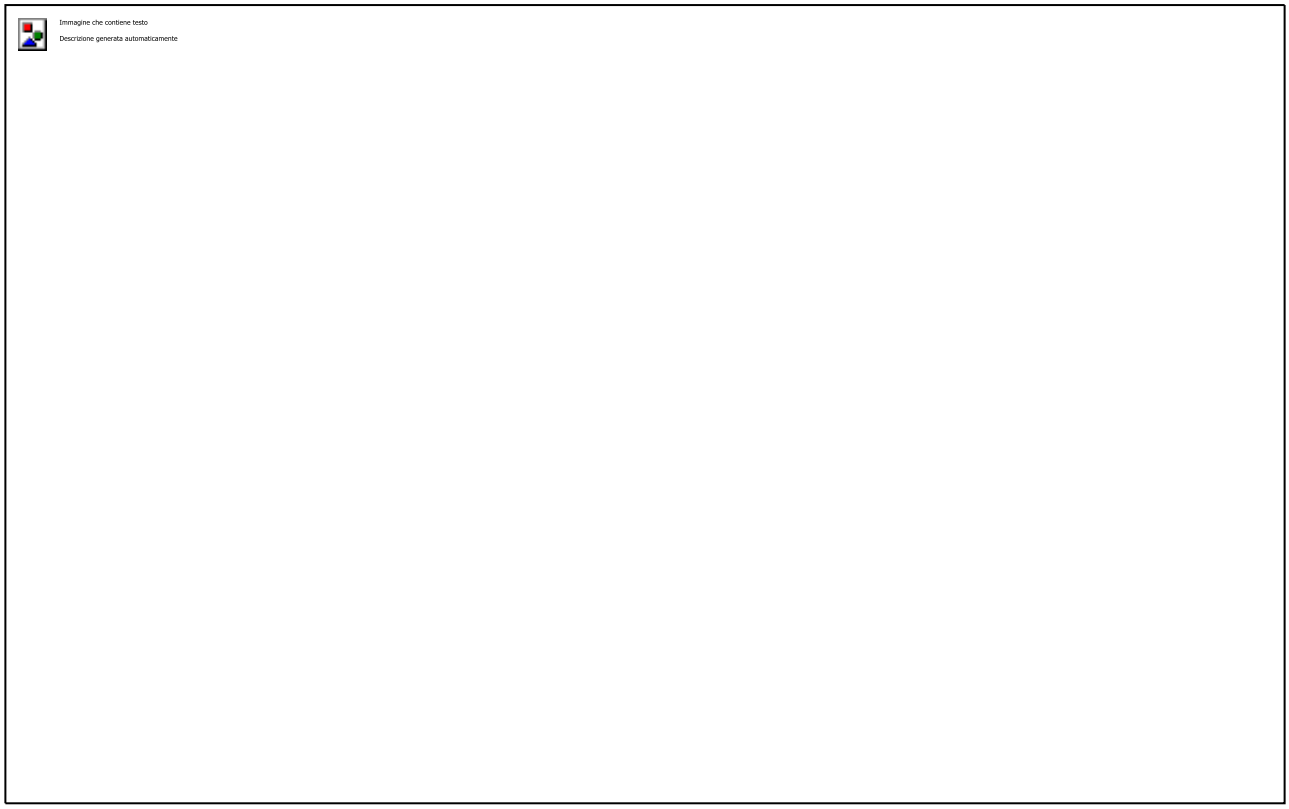


Figure A.1 – Control questions in the Dictator Game Treatment

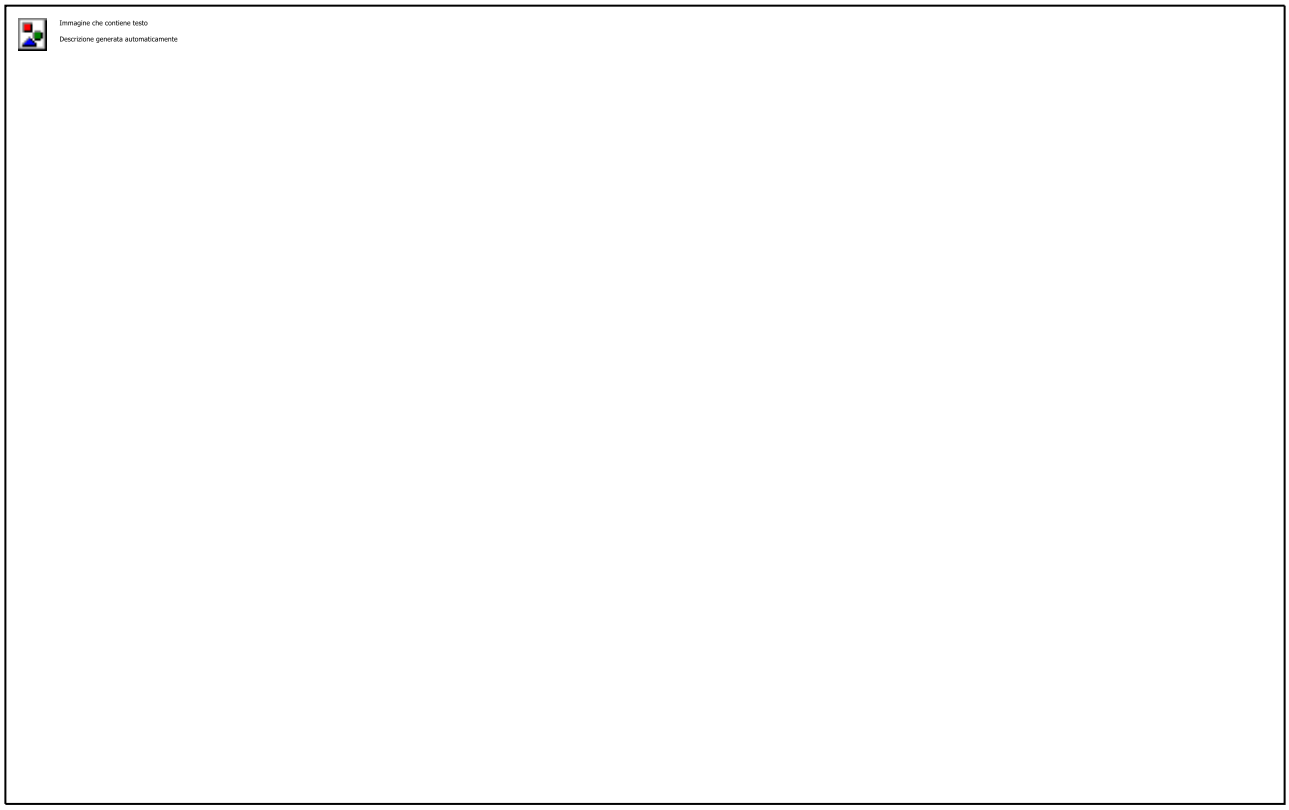


Figure A.2 – Decision screen of Participant A in the Dictator Game Treatment

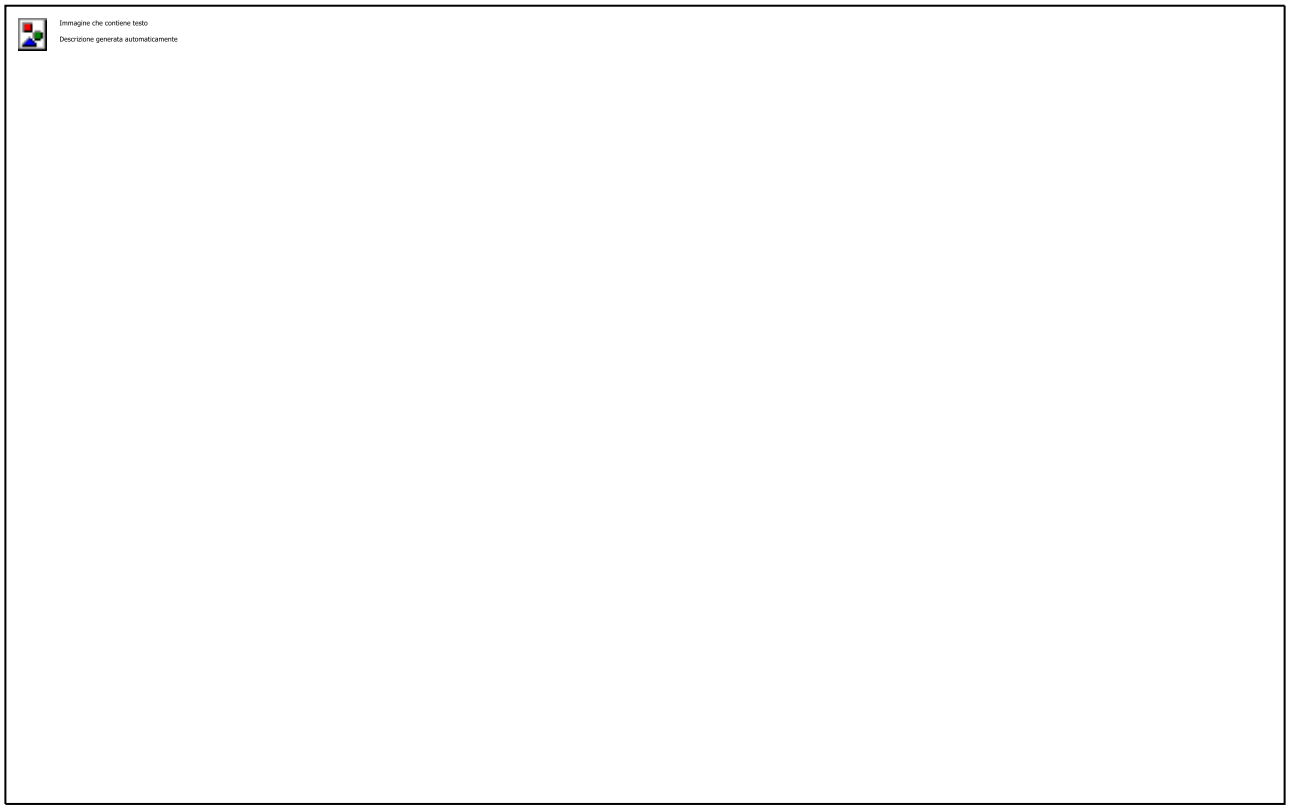


Figure A.3 – Control questions in the Third-party Punishment Game Treatment

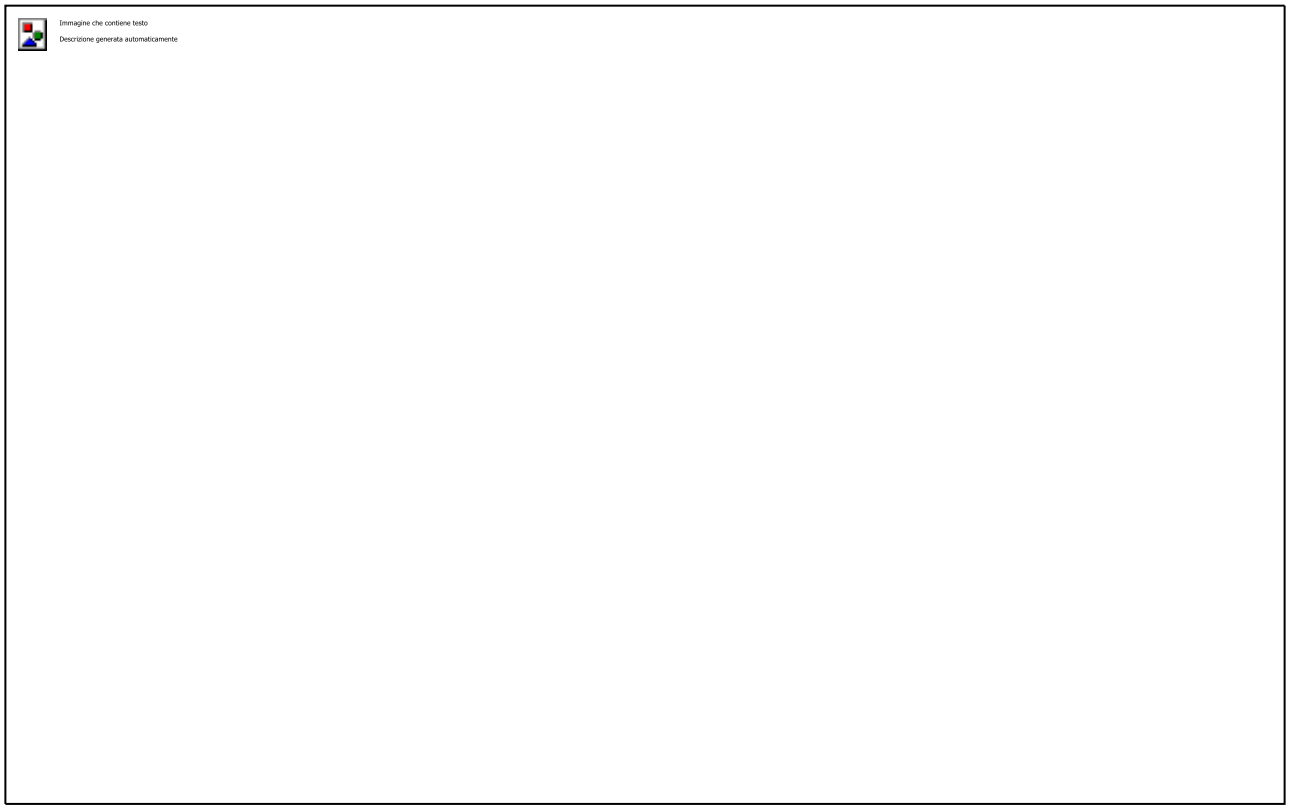


Figure A.4 – Decision screen of Participant A in the Third-party Punishment Game Treatment and in the Accountable Third-party Punishment Game Treatment

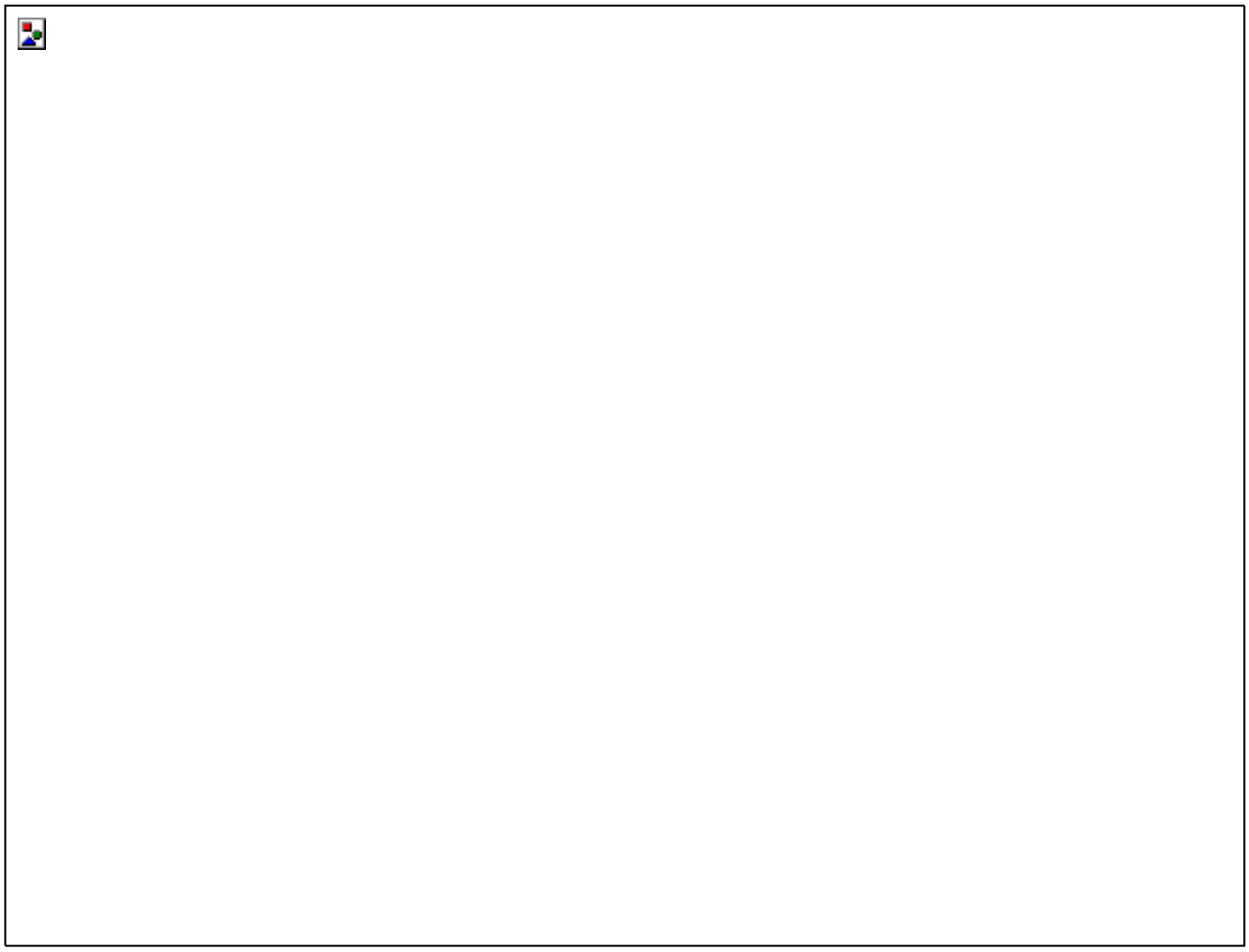


Figure A.5 – Decision screen of Participant C in the Third-party Punishment Game Treatment and in the Accountable Third-party Punishment Game Treatment

Figure A.6 – Control questions in the Accountable Third-Party Punishment Game Treatment

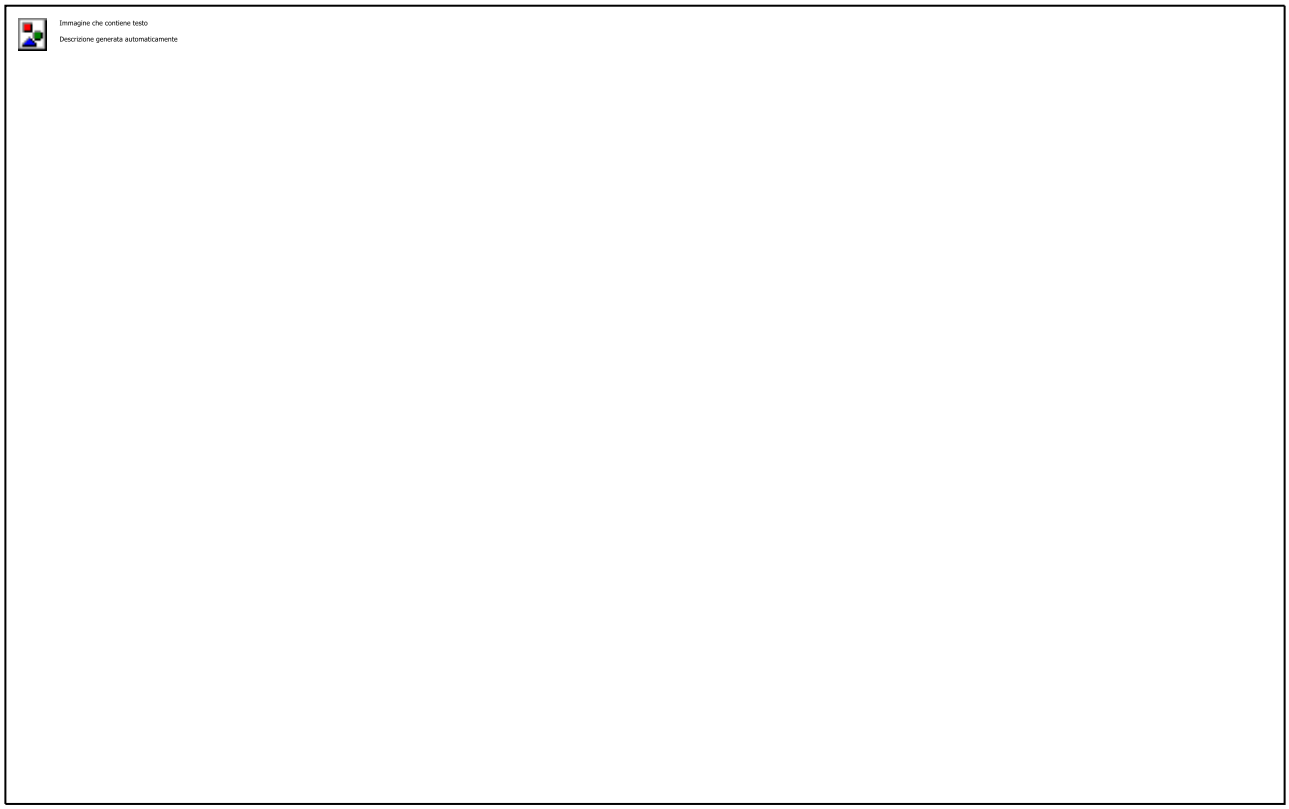


Figure A.7 – Control questions in the Accountable Third-Party Punishment Game Treatment

[1] This form of bottom-up accountability is absent whenever the judges' tenure is not appreciably affected by their actions. This holds not only for all judges in Europe, but also for all federal judges and justices of the U.S. Supreme Court. However, it is worth noting that lack of bottom-up accountability does not automatically imply independence in a broader sense. For example, Hanssen (2000) finds that U.S. states with elected justices have significantly more bureaucracies and interprets this as evidence that elected justices are more independent than appointed ones. Also see Garoupa et al. (2013).

[2] The classic contribution of Barro (1973) provides the foundations of the theory of elections as a disciplining device: by voting, citizens can limit rent extraction by elected politicians by making their reelection conditional on observed behavior.

[3] "The new politics of the internet", The Economist, January 5, 2013, pp. 16-18.

[4] All of these accountability mechanisms might be viewed as modern answers to the classic problem summed up in the famous question attributed to the Roman poet Juvenal and used here as an epigraph: *Quis custodiet ipsos custodes?* (Who watches the watchmen?)

[5] For a brief review of recent empirical work comparing accountable vs. nonaccountable officials' behavior based on field data, see Section 2.1. Unlike the studies reviewed there, Serra's (2012) paper is similar to ours in that she also runs a lab experiment including both top-down and bottom-up interventions. She finds that this combination of accountability systems turns out to be very effective in curbing misconduct. However, her focus is on fighting a specific form of wrongdoing such as corruption, so the bottom-up monitoring in her experiment, unlike in our setting, is targeted at *wrongdoers*, rather than at *third-party institutions*. For an experimental analysis of corruption centred around the reciprocity relationships emerging between bribers and public officials, see Abbink et al. (2002). For a recent field experiment showing that U.S. state legislators who are informed about being scrutinized by an independent fact-checker refrain more from making inaccurate claims, see Nyhan and Reifler (2013).

[6] In a similar vein, Lim (2013) documents that the sentencing behavior of elected justices is an important factor in determining their reelection, and that elected justices tend to be more lenient in liberal leaning districts. More broadly, as noted by Iaryczower et al. (2013), there is overwhelming evidence showing that (also outside of the U.S.) judges are sensitive to the political environment.

[7] This result is in line with previous studies on regulatory regimes focusing on electric utility bonding ratings (Formby et al., 1995), insurance regulators (Fields et al., 1997) and telephone rates (Smart, 1994).

[8] In this regard, in the aforementioned empirical investigation of electricity prices in the U.S., Besley and Coate (2003) admit that even though the states that appoint their utility commissioners and the states that elect them are similar in terms of demographic structure as well as in the proportion of electricity produced from fossil fuels, the two types of states exhibit significant differences with respect to important variables such as size, wealth and prevailing political values.

[9] In a similar vein, in the literature on the political economy of reform, Cason and Mui (2003) invite experimenters to cooperate with political economists and create different political institutions in the laboratory in order to address questions such as how differences in political institutions affect the adoption and sustainability of reform in two otherwise identical economies.

[10] Sober and Wilson (1998) report field evidence on the relevance of third-party punishment.

[11] For a recent paper based on a similar experimental design, see Almenberg et al. (2011).

[12] On the whole, then, our five-treatment experimental design allows us to compare situations where a wrongdoer acts under: 1) no third-party punishment (DG); 2) nonaccountable (or pure) third-party punishment in the presence of a single victim and of a group of victims (TP and CTP, respectively) and 3) accountable third-party punishment, which is

third-party punishment in the presence of active citizens holding third parties accountable (ATP and ACTP). See Appendix for the Instructions.

[13] That is, player A can transfer 0, 1, 2, 3, 4 or 5 tokens.

[14] The UG is the most famous example of experimental analysis of second-party sanctions.

[15] For an experimental paper on third-party punishment with multiple receivers, see Almenberg et al. (2011).

[16] This design feature aims at mapping on real-life situations where the decision taken by a person has the same impact on each subject belonging to a given category. From a methodological point of view, this is in line with our choice of all the parameters made in order to rigorously compare our results across treatments. In all treatments, each role is endowed with the same amount of money and each choice has the same monetary impact on the others. In particular, in all the treatments: a) each transfer made by player A has the same material consequence on each player B; b) each token spent by player C has the same impact on player A's payoff; c) each token spent by each player B has the same impact on player C's payoff.

[17] In law and the political sciences, recent scholarship argues that 'ideology' often underlies judicial decision-making. For an overview of this thesis, see Miles and Sunstein (2008). See also Garoupa et al.'s (2013) empirical analysis of the Spanish Constitutional Court. However, as Iaryczower et al. (2013) point out, judges' biases may reflect not only ideology but also other determinants, such as ingrained theoretical arguments about the law and personal experiences.

[18] Fehr and Fischbacher (2004) hypothesize that the salient distribution norm in the DG is the equal split, that is for A to transfer half of the 'pie' to B, arguing that since the players interact anonymously and are randomly assigned their roles, there is no reason why A should end up with more money than B. They also elicit fairness judgments that clearly indicate that the egalitarian solution is perceived to be the fair solution. Bernhard et al.'s (2006) experimental finding suggests the existence of an egalitarian sharing norm. However, the fact that subjects' sense of fairness may be context-dependent was clear for example in Ottone (2008). When participants have to *earn* their endowment, the Observers both punish and transfer *less* than when the endowment is randomly assigned. This suggests that the fairness reference point changes as the situation changes. List's (2007) study, based on the DG framework, confirms that this is the case, reporting experimental evidence from nearly 200 dictators in treatments that varied the action set and the origin of endowment (earned vs. unearned). In particular, he shows that many fewer subjects are willing to transfer money when the action set includes taking. Henrich et al. (2001) find that what is unfair in one society is perfectly fair in another one.

[19] Bernhard et al. (2006), in their third-party punishment experiments, elicit players' empirical expectations (but not their normative ones) about how the dictators should be punished at different transfer levels.

[20] The pioneering studies using the notion of a reference point for fairness in formal models of distributive preferences are Falk and Fischbacher (2000) and Cox et al. (2007). For a recent axiomatization of reference-dependent preference structures closely related to well-known experimental works dealing with fairness-driven behavior, see Sandbu (2008).

[21] In this regard, the DG treatment provides us with a clean test, as it allows us to elicit participants' fairness reference point in the absence of Third Parties who might intervene. Subsequently, by moving from the DG to the TP treatment, we are able to see whether the presence of a third-party institution that can implement a sanctioning mechanism changes subjects' ideas of what is fair. Hence, within our experiment as a whole, the DG treatment serves the purpose of providing a reference point for evaluating Third Parties' behavior in the TP.

[22] This methodological choice owes to the fact that we did not want to focus players' attention on fairness issues before they made their choice. Details concerning the elicitation procedure are found in Section 5.1.

[23] Translation of instructions is in the Appendix (the original text in Italian is available upon request).

[24] The main advantage of this method is that it significantly improves statistical power as it provides information about responses to outcomes that may occur very infrequently (see on this also Fehr and Fischbacher, 2004). The strategy method allows the researcher to analyze sanctioning behavior for *each* possible transfer level from the Dictator to the Receiver: since Dictators rarely choose certain transfer levels, if C could respond to A's *actual* choice only, it would be impossible to perform this analysis. Several tests in simple games have not found behavior induced by the strategy method to be significantly different from that induced by the standard direct-response method (Charness et al., 2008). Also in Henrich et al. (2006) the strategy method was used with regard to third-party punishers' decisions.

[25] Therefore, in all five treatments we also expect that, if we assume that common knowledge holds and that, hence, the Dictator believes that the Third Party and (in the ATP and in the ACTP) the Receiver are selfish (so the threat of their punishing is not considered credible), a selfish Dictator transfers nothing to the Receiver. As a consequence, under the above assumptions, the following subgame perfect equilibrium outcomes are predicted, in the five treatments: zero transfer by the Dictator in the DG; zero transfer by the Dictator and zero punishment by the Third Party in the TP and in the CTP and zero transfer by the Dictator and zero punishment by both the Third Party and the Receiver in the ATP and in the ACTP.

[26] This hypothesis also presupposes that the so-called *Homo Oeconomicus* model does not adequately capture human behavior in strategic interaction contexts such as the ones investigated in this work.

[27]Notice that, when the Dictator transfers ¼ of her endowment to the Receiver, both players end up with 15 tokens each. Consequently, ¼ is the transfer that would be in line with the 50/50 norm.

[28] We perform t-tests on the pooled sample since a series of Kruskall-Wallis tests confirms that there is no significant difference of the fair transfer within the same role along the treatments (see Table 3).

[29] The same tests are performed on both the Dictators and the Receivers. The results are the same.

[30] The same is true if we compare the average fair transfers according to both the Receivers and the Dictators along the treatments (Kruskall –Wallis test, $p = 0.24$ and $p = 0.88$ respectively).

[31] When we compare different participants' normative beliefs to check whether the role they are called upon to play is relevant in determining people's perception of fairness, it turns out that the Receiver's fair transfer is significantly higher than those of the Dictators' and the Third Parties' (t-test, $p = 0.000$ in both cases). On the other hand, Dictators' and Third Parties' beliefs are aligned (ttest, $p = 0.919$). However, we find that, regardless of their role, the players share the idea that the fair transfer from the Dictator to the Receiver is significantly different from both 0 (the selfish choice) and 5 (the egalitarian choice). This suggests that stakeholders are not impartial in their judgments about 'what is fair'. This finding is compatible with the intuitive argument that citizens are characterized by a *self-serving bias in moral reasoning* (see on this Babcock and Loewenstein, 1997, and Konow, 2001). Further, it seems to be in line with a famous sentence by Milton Friedman: "'Fairness' is strictly in the eye of the beholder": to Receivers, a 'fair' transfer is a relatively high transfer, compared to Dictators and Third Parties. However, since we also find that what we might call 'selfishness aversion' emerges across roles and across treatments, we can also say that not all fairness seems to be strictly in the eye of the beholder.

[32] For example, if the Third Party declares that the Dictator ought to transfer 3 tokens to the Receiver, we can observe the Third Party's reaction to -3, -2, -1, 0, 1 and 2 levels of subjective unfairness.

[33] In this regard, the media play a potentially important role, within modern democracies. Lim et al. (2012) report that the vast majority of voters in the U.S. say that they have insufficient information about judicial candidates. Further, the media only occasionally provide this information, with the amount of available press coverage varying enormously, from none to hundreds of articles per newspaper, judge and year in their sample. Hence, it is reasonable to argue that the functioning of alternative selection methods of public officials is heavily dependent on variations in media coverage.

[34] Moreover, the new setting will allow us to carry out a welfare analysis. A peculiarity of sanctioning mechanisms is that resources are destroyed in the short run in order to enforce norms and have advantages in the long run – when norms are enforced, no violation occurs and punishment is a credible threat that need no longer be implemented. This implies that in a one-shot game framework, punishment may generate a welfare loss. However, since we found that people are ready to sanction unfair behavior within a one-shot game protocol where strategic considerations play a limited role, there is no reason to conjecture that subjects will abstain from issuing sanctions in a repeated game context,

where there is far more room for both nonstrategic and strategic motives for punishment. Hence, we hypothesize that our results will be confirmed and reinforced in the new setting. Our intuition is that, in the long run, punishment will act as a credible threat and that, consequently, this would discourage norm violation and therefore improve social welfare.

References

Abbink, K., B. Irlenbusch, and E. Renner. 2002. "An experimental bribery game" 18 Journal of Law, Economics, and Organization 428-454.

Alesina, A., and G. Tabellini. 2007. "Bureaucrats or politicians? Part I: A single policy task." 97 American Economic Review 169-179.

Almenberg, J., A. Dreber, C.L. Apicella, and, D.G Rand, 2011. Third party reward and punishment: group size, efficiency and public goods, in Palmetti N.M. et al. (eds.), *Psychology of Punishment*, Nova Science Publishers.

Babcock, L., and G. Loewenstein. 1997. "Explaining bargaining impasse: the role of self-serving biases," 11 Journal of Economic Perspectives 109-126.

Barro, R.J.. 1973. "The control of politicians: an economic model," 14 Public Choice 19-42.

Bernhard, H., E. Fehr, and U. Fischbacher. 2006. "Group affiliation and altruistic norm enforcement," 96 American Economic Review 217-221.

Bernhard, H., U. Fischbacher, and E. Fehr. 2006. “Parochial altruism in humans,” 442 Nature 912-915.

Besley, T., and A. Case. 2003. “Political institutions and policy choice: evidence from the United States,” 41 Journal of Economic Literature 7-73.

Besley, T., and S. Coate. 2003. “Elected versus appointed regulators: theory and evidence,” 1 Journal of the European Economic Association 117-1206.

Besley, T., and A. Payne. 2013. “Implementation of anti-discrimination policy: does judicial selection matter?” 15 American Law and Economics Review 212-251.

Bolton, G.E., and A. Ockenfels. 2000. “A theory of equity, reciprocity and competition,” 90 American Economic Review 166-93.

Carpenter, J., and P.H. Matthews. 2009. “What norms trigger punishment?” 12 Experimental Economics 272-288.

Carpenter, J., and P.H. Matthews. 2010. “Norm enforcement: the role of third parties” 166 Journal of Institutional and Theoretical Economics 239-258.

Cason, T.N., and V. Mui. 2003. “Testing political economy models of reform in the laboratory,” 93 American Economic Review 208-212.

Chang, E., M. Golden, and S.J. Hill. 2010. “Legislative malfeasance and political accountability,” 62 World Politics 177-220.

Charness, G., R. Cobo-Reyes, and N. Jimenez. 2008. “An investment game with third-party intervention,” 68 Journal of Economic Behavior and Organization 18-28.

Choi, S.J., G.M. Gulati, and E.A. Posner. 2010. “Professionals or politicians : the uncertain empirical case for an elected rather than appointed judiciary,” 26 Journal of Law, Economics, and Organization 290-336.

Cox, S.H., L.G. Epstein, and S. Gjerstad. 2007. “A tractable model of reciprocity and fairness,” 59 Games and Economic Behavior 17-45.

Falk, A., E. Fehr,, and U. Fischbacher. 2008. “Testing theories of fairness – Intentions matter,” 62 Games and Economic Behavior 287-303.

Falk, A., and U. Fischbacher. 2000. “A theory of reciprocity,” 54 Games and Economic Behavior 293-315.

Fehr, E., and U. Fischbacher,. 2004. “Third-party punishment and social norms,” 25 Evolution and Human Behavior 63-87.

Fehr, E., and S. Gächter. 2000. “Cooperation and punishment,” 90 American Economic Review 980-994.

Fehr, E., and K.M. Schmidt. 1999. “A theory of fairness, competition and cooperation,” 114 Quarterly Journal of Economics 817-51.

Fields, J.A., L.S. Klein, and J.M. Sfiridis. 1997. “A market based evaluation of the election versus appointment of regulatory commissioners,” 92 Public Choice 337-351.

Fischbacher, U.. 2007. “z-Tree: Zurich toolbox for ready-made economic experiments,” 10 Experimental Economics 171-178.

Formby, J.P., B. Mishra, and P.D. Thistle. 1995. “Public utility regulation and bond ratings,” 84 Public Choice 119-136.

Forsythe, R., J.L.Horowitz., N.E. Savin, and M. Sefton. 1994. “Fairness in simple bargaining experiments,” 6 Games and Economic Behavior 347-369.

Garoupa, N., F. Gomez-Pomar, and V. Grembi, 2013. “Judging under political pressure: an empirical analysis of constitutional review voting in the Spanish constitutional court,” 29 Journal of Law, Economics, and Organization 513-534.

Gordon, S., and G. Huber. 2007. “The effect of electoral competitiveness on incumbent behaviour,” 2 Quarterly Journal of Political Science 107-138.

Guiso, L., P. Sapienza, and L. Zingales. 2011. Civic capital as the missing link, in J. Benhabib, A. Bisin, and M. Jackson, M. (eds.), *Handbook of Social Economics*, Elsevier.

Güth, W., R. Schmittberger, and B. Schwartze. 1982. „An experimental analysis of ultimatum games,” 3 Journal of Economic Behavior and Organization 367-388.

Hanssen, F.A.. 1999. “The effect of judicial institutions on uncertainty and the rate of litigation: the election versus appointment of state judges,” 28 Journal of Legal Studies 205-232.

Hanssen, F.A.. 2000. “Independent courts and administrative agencies: an empirical analysis of the states,” 16 Journal of Law, Economics, and Organization 534-571.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, and R. McElreath. 2001. “Cooperation, reciprocity and punishment in fifteen small-scale societies,” 91 American Economic Review 73-78.

Henrich, J., R. McElreath, A. Barr, and J. Ensminger. 2006. “Costly punishment across human societies,” 312 Science 1767-1770.

Iaryczower, M., G. Lewis, and M. Shum. 2013. “To elect or to appoint? Bias, information, and responsiveness of bureaucrats and politicians,” 97 Journal of Public Economics 230-244.

Konow, J.. 2001. “Fair and square: the four sides of distributive justice,” 46 Journal of Economic Behavior and Organization 137-164.

Levine, D.. 1998. “Modelling altruism and spitefulness in experiments,” 1 Review of Economics Dynamics 593-622.

Lewisch, P.G., S. Ottone, and F. Ponzano. 2011. “Free-riding on altruistic punishment? An experimental comparison of third-party punishment in a stand-alone and in an in-group environment,” 7 Review of Law and Economics 161-190.

Lim, C.S.H.. 2013. “Preferences and incentives of appointed and elected public officials: evidence from state trial court judges,” 103 American Economic Review 1360-1397.

Lim, C.S.H., and J.M. Snyder. 2012. “The judge, the politician and the press: newspaper coverage and criminal sentencing across electoral systems,” mimeo, Cornell University.

List, J.. 2007. “On the interpretation of giving in dictator games,” 115 Journal of Political Economy 482-493.

Marlowe, F.W., J.C. Berbesque, A. Barr, C. Barrett, A. Bolyanatz, J.C.Cardenas, J. Ensminger, M. Gurven, E. Gwako, J. Henrich, N. Henrich, C. Lesorogol, R. McElreath, and D. Tracer. 2008. “More ‘altruistic’ punishment in larger societies,” 275 Proceedings of the Royal Society Biology 587-590.

Maskin, E., and J. Tirole. 2004. “The politician and the judge: accountability in government,” 94 American Economic Review 1034-1054.

Miles, T.J., and C.R. Sunstein. 2008. “The new legal realism,” 75 University of Chicago Law Review 831-852.

Nannicini, T., A. Stella, G. Tabellini, and U. Troiano. 2013. “Social capital and political accountability,” 5 American Economic Journal: Economic Policy 222-250.

Nyhan, B., and J. Reifler. 2013. “The effects of fact-checking threat: results from a field experiment in the states,” New America Foundation Media Policy Initiative Research Paper.

Ottone, S.. 2005. “Transfers and altruistic punishments in Solomon’s game experiments,” Paper n. 57, AL.EX Series, POLIS, University of Eastern Piedmont.

Ottone, S.. 2008. “Are people Samaritans or avengers?” 3 Economics Bulletin 1-8.

Peters, J.G., and S. Welch S. 1980. “The effects of charges of corruption on voting behavior in congressional elections,” 74 American Political Science Review 697-708.

Pinello, D.R.. 1995. *The impact of judicial-selection method on state-supreme-court policy: innovation, reaction, and atrophy*. Westport, CT: Greenwood Press.

Rabin, M.. 1993. “Incorporating fairness into game theory and economics,” 83 American Economic Review 1281-1302.

Reinikka, R., and J. Svensson. 2004. “Local capture: Evidence from a central government transfer program in Uganda,” 119 Quarterly Journal of Economics 679-705.

Sandbu, M.E.. 2008. “Axiomatic foundations for fairness-motivated preferences,” 31 Social Choice and Welfare 589-619.

Serra, D.. 2012. Combining top-down and bottom-up accountability: evidence from a bribery experiment. 28 Journal of Law, Economics, and Organization 569-587.

Smart, S.R.. 1994. “The consequences of appointment methods and party control for telecommunications pricing, 3 Journal of Economics and Management Strategy 301-323.

Sober, E., and D.S. Wilson 1998. *Unto others: the evolution and psychology of unselfish behavior*, Cambridge (Mass.): Harvard University Press.

Takahashi, N., M. Miyahara, and R. Mashima. 2005. “Does intention matter in third-party punishment?” Paper presented at the 17th annual meeting of Human Behavior and Evolution Society, Austin, Texas, June 1-5.