EUI WORKING PAPERS

SPS 2015/01
Department of Political and Social Sciences

Fairness and Compliance in the Extortion Game

Giulia Andrighetto, Daniela Grieco and Rosaria Conte

European University Institute
Department of Political and Social Sciences

Fairness and Compliance in the Extortion Game

Giulia Andrighetto[1,2], Daniela Grieco[3] and Rosaria Conte[2]

[1]European University Institute, Fiesole

[2]Italian National Research Council, Institute of Cognitive Science and Technology, Rome

[3]Bocconi University, Department of Economics, Milan

EUI Working Paper **SPS** 2015/01

ISSN 1725-6755

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu
cadmus.eui.eu

Abstract

In this paper we introduce the "extortion game" and propose a set of experiments aimed at studying both the extorter's and the victim's behavior. The primary focus of our experiments is understanding what determines both the extent of the extortive request and the victim's reaction in terms of compliance. Our results show that the extorter's requests are proportional to the victim's earnings, similar across victims and positively correlated with the level of request the extorter perceives as "fair". Punishment is rare and the extorted sums are rather small. Results also shed light on the motivations that make a victim compliant with extortive requests, confirming that punishment plays a role in promoting compliance, but also showing the importance of other concurrent factors like information about peers' behavior and the perceived fairness of the request. Overall, our results indicate that fairness concerns matter even in a context of extra-legal taxation, shaping both extorters' requests and victims' compliance

Keywords

Extortion, social norms, fairness, compliance, inequity aversion, laboratory experiments.

Introduction[1]

The present work aims at understanding the reasons why people do comply to an appropriative request and the mechanisms at the basis of victims' compliance. We introduce a new game, called "extortion game", that has been designed specifically to investigate repeated interaction among subjects where one of them can appropriate part of the others' earnings. The design presents a series of features that have never been investigated in the same game: we use a "take" frame instead of the typical "give frame" used in ultimatum, dictator and public goods games usually studied to explore fairness, equity and reciprocity. Additionally, we manipulate asymmetry in subjects roles and in the possibility to earn profits, and we allow for the presence of multiple potential victims of appropriation, information sharing between them, and multiple rounds in order to explore how extortion and compliance evolve in time.

Our setup introduces crucial elements of novelty that, first, enable us to derive further and original insights on fairness with respect to experimental evidence in the existing literature. Second, we propose that the attributes characterizing an extortive request and the cognitive processes leading victims to comply with it could present some similarities with those at the basis of norms' compliance. We will then apply the growing knowledge in the domain of norm compliance to understand behaviors in the extortion game so to identify a possible common core between an appropriative and a normative request.

Standard theories of norm compliance advocating "punishment, as the most effective way of making people compliant with norms" (Coleman, 1990) have been largely challenged (Bicchieri, 2006; Fehr & Fischbacher, 2004, Conte at al. 2014, Gintis, 2000) and a set of other factors have been proven relevant in motivating people to obey norms, such as the legitimacy and fairness of the request, the conformity and normative expectations of others, and the salience of the norm in question (Andrighetto et al. 2013; Bicchieri, 2006; Tyler, 2006; Xiao & Houser, 2011). Similarly, obedience to an extortive request has been mainly explained as the result of threats, coercion and violence: "the extortive activity is effective because its

[1] The work presented in this paper has been performed in the frame of the GLODERS (FP7/2007-2013) project. The authors thank Jordi Brands, Daniela Di Cagno, Diego Gambetta, Alessandro Innocenti, LucaTummolini, the partners of the GLODERS project and the participants in the ESA World Meeting in Zurich, the EUROCRIM Conference in Budapest, the 8th Alhambra Experimental Workshop in Rome and the 2014 SEET Workshop in Sesimbra for insightful comments regarding earlier versions of this paper.

victims know in advance that there is the possibility to suffer violent retaliations when the request is not accepted" (La Spina, 2008; Becker, 1968). An explanation of compliance to extortive requests exclusively in terms of violence and coercion risks to underestimate other important motivating factors and then inspire measures for contrasting the extortive phenomenon only partially effective.

Our analysis investigates both the extorter's and the victim's behavior: the primary focus of our experiments is understanding which factors make an extortive request a persuasive and what motivates a victim to comply with it. The focus is on how victims adjust their behavior depending on a number of factors, namely the level of punishment, the perceived fairness of the received requests, and the information about the behavior of others. Among the factors that compliance to extortion might be traced back to, these have been addressed as they are have been shown also to affect norm compliance (Fehr & Gächter, 2002; Fischbacher, Gächter, and Fehr, 2001; Bazart and Bonein, 2014; Cialdini, Kallgren, and Reno, 1991; Ostrom, Walker and Gardner, 1992; Tyler, 2006). The underlying intuition is that some of the factors leading to norm compliance might also affect the victims' reactions to extortive requests. Fairness concerns matter even in a context of extra-legal taxation, affecting both the extorters and victims' decisions. After recalling the existing literature that can be related to this work (Section 2), we present our research questions (Section 3). In Section 4 we describe the "extortion game". Results described in Section 5 show that: on the extorter's side: (I) the extorter requests are balanced to the victim's earnings and similar across the different victims, showing that each extorter has his peculiar "extortive style", i.e. he can be more or less predatory with his victims but in this he is consistent across potential victims; (II) extorters who choose their role make lower requests compared to those who were randomly assigned to this role, showing some sense of "responsibility" and fairness towards their victim (with respect to the Extorter who gets the role randomly); (III) requests are significantly lower when the Extorter has his own source of earnings; (IV) punishment is rare but plays a crucial role in increasing compliance; on the victim's side: (V) information about others' compliance are shown to play a key role in promoting victims' compliance, and (VI) the *perceived fairness* of the extortive request is significantly and positively correlated with the request made by the extorter and shapes victims' compliance with the request (the less the request is perceived as fair by the victim, the lower the level of compliance). In Section 6, we discuss these findings and conclude.

2. Related literature

A broad set of experimental findings show that people frequently make decisions that do not maximize their monetary payoffs. Experimental subjects reject positive offers in the ultimatum game, allocate positive amounts to anonymous recipients in the dictator game, contribute voluntarily to public good games although the marginal private return of free-riding would be higher (Charness and Haruvy, 2002). In the last two decades experimental economists have gathered overwhelming evidence that people are strongly motivated by concerns for fairness. Games like the Ultimatum Game, the Dictator Game, the Gift Exchange Game, the Trust Game, Public Good and Common Pool Resources Games are now well-established and – although from different angles – all provide clear evidence that people make costly decisions in order to behave fairly and to sanction peers' unfair behavior. Besides the experimental outcomes, fairness considerations are relevant in several domains of everyday life in individuals' interaction with family, friends, and even strangers, but also shape the behavior of people in important economic domains like firms' policies, procedural justice, collective action problems, and management of shared resources (Fehr and Schmidt, 2003).

Since the game we propose makes use of a "take" frame, it is worthwhile recalling the key aspects of the appropriation game as discussed by Cox and al. (2013) in its fairness concerns. Games reproducing appropriation from a common-pool resource ground on the seminal work by Walker et al. (1990) where a group of players take part into a non-cooperative game in which each player makes an appropriative decision: in particular, players allocate an endowment between a Common Pool Resource and a private alternative. The theoretical prediction is over-appropriation, in the same extent under-provision is expected in voluntary contribution mechanism game, since the only difference for a selfish player is the framing ("take" versus "give") of the two games. Experimental results show that, while symmetric provision and appropriation games produces comparable behaviors, power asymmetry leads to significantly higher appropriation game than in a payoff-equivalent provision game because second-movers generally react by reciprocating previous altruist or selfish behavior: when the opponents' decision is observable, subjects with stronger power appear to be more sensitive towards opponents' selfish choices and react by destroying surplus in order to sanction a level of appropriation they judge as "unfair".

The "power-to-take" game (Van Winden, 2001; Rueben and Van Winden, 2010) is an interesting version of the "ultimatum game" with the "take" frame. In this game, a proposer can make a claim on the resources of the responder; then, the responder can refuse the proposal by destroying any part of her own resources (zero and the whole endowment are possible options). In this game, each participant has an endowment, but only the responder's endowment is at stake; the responder can destroy any amount of her own endowment. Traditional economic theory predicts that a proposer will choose to take all the responder's endowment and that the responder will not destroy: however, responders typically destroy some or all of their endowment when faced with requests perceived as "unfair" (i.e., requests that are higher than 50% of the endowment), and the proposer hardly asks for all the responder's endowment. This game can be used to capture agency situations where a principal decides on an incentives' scheme for an agent, and the agent's effort can be crowded out, thus reducing both his own and the principal's payoff: if an employer is too demanding, a worker could decide to reduce effort; if a Government asks too much in terms of taxes, a worker could reduce his working hours, and so on. The "power to take" game differs from the "the ultimatum-game" (e.g. Guth, Schmittberger and Schwarze, 2003; Camerer, 2003) in the fact that in the latter the proposer makes a claim on how to split her own endowment with the responder ("give" frame), and the responder may accept or not. Results show that the offers are significantly higher than zero, and that responders refuse unfair and unequal offers although this is not economically convenient for them. Both power-to-take and ultimatum games are interesting for studying a dyadic interaction where either the request or the offer (respectively) turn out to be motivated by reasons that account for both "profit maximizing" (proposers do not want their request or offer to be rejected), but also for "nonprofit-maximizing" behavior, i.e., are driven by fairness and inequality aversion considerations.

Another game, although not tested experimentally, present elements related to extortion, i.e the "blackmail game" (Ellsberg, 1968). As quoted in Ellsberg's paper, "the term blackmail first entered the English language in the 1500s. It referred to tribute exacted by families or clans along the Scottish-English border in return for immunity from raids by Scottish or English bands. One way to make a living during that era was to steal sheep and horses from the English or Scots and others not paying *protection money*." In this game, the (potential) victim of blackmailing has an endowment that she can agree to split with the blackmailer: the amount to be given to the blackmailer is exogenous and fixed. The victim can accept or

refuse, facing the chance of being victim of retaliation. Punishment is depicted as exceptionally dangerous for both parties: the victim lose her whole endowment, and the blackmailer becomes visible to concurrent bands who aim at controlling the land: both players' payoff drops to zero. This game has not been tested experimentally, but the theoretical prediction is observing both no compliance and no punishment, as it is extremely costly for the blackmailer to carry out the threatened punishment. Ellsberg admits that observing compliance in the real world is plausible: a victim estimates the probability to be punished, which could be more or less uncertain according to the specific situation, and decides whether or not to yield based on her level of "critical risk": to yield, the victim must believe that the likelihood of punishment is higher than her own critical risk. Ellsberg also considers the chance to modify the payoffs in order to account for binding commitment: if the blackmailer *does have* to punish, the threat needs to be made plausible. The payoffs of the blackmailer are lowered in case the blackmailer does not punish, in order to capture the losses in terms of honor, prestige and reputation that derive from neglecting the deterrent role of a credible threat. In the game, this can be obtained by irreversible destruction of alternatives, or by assigning the right to punish to a third party.

Similarly, predatory-prey models à la Grossman and Kim (1996) capture the interaction of a potential predator who chooses between production and appropriation and a prey who chooses between production and defensive fortification to be protected against appropriation. Both players hold an initial resource endowment, but only defense's endowment is subject to seizure. Carter and Anderton (2001) test the predatory-prey model in the lab introducing three treatments that vary the effectiveness of predation from low to medium to high. Although the experiment payoff structure circumscribes the opportunities to exercise fairness concern, the experimental results show that the convergence to the sub-game perfection is reinforced by preference for an equal distribution of payoffs.

3. Research questions

The first set of research questions the "extortion game" deals with the extorter's *request* and in particular with how he decides to shape it in order to make it more persuasive: is the extortive request tailored on the victim's level of profits or independent from them? Is the request predatory or reasonable, let's say "fair"? If an extortive style can be identified, is it applied to all victims? Is the request affected by the way the extorter acquired this role?

The second set of questions is related to victims' *compliance* and is aimed at understanding why victims obey to extortive requests: is it just because of the fear of punishment, or is there any other concurrent factor contributing to its compliance? Does the fact that an extortive request is perceived as "fair" play any role in shaping the victims' compliance? Is compliance conditional to the behavior of other victims or not?

4. Experimental design

Participants are divided in groups of three subjects, where one subject is assigned to the role A and the remaining two subjects are assigned to the role B. In the instructions, we avoided the use of any loaded terms and labeled with A the extorter and with B the potential victims of extortion.

Subjects in each group interact for T=10 periods: each period is composed by four stages (see below for a detailed description). The features of interaction that are common across treatments are the following: subjects of type A can make a request over subjects of type B's earnings: extortion occurs once B (eventually after bargaining) accepts the request. Consequently, A's appropriation of the requested amount of B's tokens takes place. Furthermore, each A has the possibility to punish B by affecting her future chances to earn money.

Concerning the extortion choice, A has two contrasting objectives: appropriate the more he can, but without crowding out B's effort. Concerning the punishment choice, punishment is costly, as it reduces A's chances of further appropriation, but gives A the chance to send a signal to a non-compliant B (sanctioning or signaling role of punishment) or to prevent his decision to be non-compliant (deterrent role of punishment). Although we expect A to use punishment only against non-compliant Bs, there is no formal restriction on the punishment target (in principle, also compliant Bs can be punished).

The theoretic prediction for the one-shot game is the following: since punishment is costly not only for B (as it decreases the chance to earn tokens) but also for A, A should never punish B. As B anticipates a zero profit ($p_A=0$), then she should decide to be non-compliant ($g_B=0$). As A anticipates $g_B=0$, then she should decide to make no request ($r_A=0$): extortion should not exist. This game leads to a conclusion that is opposite to the prediction for the "power-to-

take" game, where the Responder should accept from the Proposer any request but 100%, and the Proposer should request (100-ε)%. Furthermore, it presents a multi-lateral interaction that is useful to investigate how multiple receivers of an extortive request may influence each other in terms of the level of compliance chosen.

4.1 Treatments

In this section, a description of the four treatments of the "Extortion game" is provided. See Table 1 for an overview of the treatments.

[TABLE 1]

Baseline (Treatment 1)

In the first stage, all the subjects earn their endowment by taking part in a real-effort task consisting of 3 general knowledge multiple-choice questions with 4 possible answers (where only one is correct). They have 25 seconds to answer each question. Earnings are expressed in tokens and depend on the number of correct answers (40 tokens for each correct answer) and how fast they are (each second saved is worth one token). A has the chance to earn her endowment in the same way. We choose to use a task which requires both effort and ability to explore whether any of them are influenced by the possibility of extortive request and whether effort and/or ability are crowded out if a subject is victim of appropriation.

In the second stage, A receives a feedback on the amount π_B earned by each B in her group and decides how much to appropriate[2] (r_A) out of π_e: also zero and the whole earning π_B are possible choices. The ratio r_A / π_B captures A's proportional request (when proportion is set with respect to B's earnings); the comparison between r_A / π_B and r_A / π_{-B} could be used as a proxy for A's fairness in treating the two potential victims.

[2] Again, we do not use loaded terms: instructions speak of "taking" and not of "extorting".

In the third stage, B receives A's request and decides how much to give (g_B) out of r_A: also zero and the whole request r_A are possible choices. The ratio g_B/r_A captures the degree of B's compliance.

In the fourth stage, A receives a feedback on g_B and has the chance to punish each B in her group, by eliminating one or more (up to all the three) questions. A has three options: eliminating one question at the cost of 10 tokens (that we *ex post* classify as "low-intensity punishment"), eliminating two questions at the cost of 20 tokens ("medium-intensity punishment"), eliminating three questions at the cost of 30 tokens ("high-intensity punishment"). In other words, A can choose between 1, 2 or 3 punishment points.

Role choice (Treatment 2)

This treatment differs from the Baseline in the fact that the roles of A or B are assigned after subjects express their preference for assuming the role of A by answering to the following statement: "Please express your preference for assuming the "A" role instead of the "B" role from 1 (no desire at all) to 10 (very strong desire)". The subjects make this decision after being aware of all the relevant features of both roles, and before starting the ten rounds of interaction.

Role choice + asymmetric endowment (Treatment 3)

This treatment differs from the previous in the fact that *only* subjects of type B can earn their money by answering questions; subjects of type A can make money only by appropriating subjects B's earnings.

Feedback (Treatment 4)

This treatment differs from Treatment 3 in the fact that subjects of type B receive a feedback on the level of compliance of the other *B* in their group in the previous period. The feedback is reported as follows: "In the previous round, the other B subject in your group has given *X* tokens out of the *Y* tokens A has requested".

4.2. Procedures

The experiments were conducted at the CESARE Lab of LUISS University in Rome. Subjects were recruited *via* ORSEE (Greiner, 2004). We ran 8 computerized[3] sessions between June 2013 and July 2013, with a total of 171 participants (45 subjects in Treatments 1, 2 and 4; 36 subjects in Treatment 3). Participants were undergraduate students (63.2% from Economics), with 57.3% males. We employed a between-subjects design: no individual participated in more than one session. In each session, the participants were paid a 2€ show-up fee, plus their earnings from the experiment. At the beginning of each session, participants were welcomed and, once all of them were seated, the instructions were handed to them in written form before being read aloud by one experimenter. All subjects completed a final questionnaire containing demographic information, a statement about the level of percentage request they judged as "fair", and a set of 16 questions where they had to self-report (in a 1-7 scale) the intensity of the emotions they felt during the experiment. The sessions took approximately 45 minutes, with earnings ranging between 2€ and 19€.

5. Results

5.1. Descriptive statistics

With the exception of the Baseline Treatment, the attribution of experimental role is not random, but based on a preference for playing the role of A instead of B expressed by all subjects at the beginning of the experiment. As anticipated above, this preference is expressed in a scale from 1 to 10, where 10 is the maximum willingness to assume the A role. Table 2 summarizes the average level of preference expressed by subjects in the four treatments.

[TABLE 2]

[3] The experiment was programmed by using the z-tree platform (Fischbacher, 2007).

The preference for A's role is significantly lower when As have no way to earn money other than taking tokens from their Bs (what we call "asymmetric endowment"): the average preference expressed in Treatment 2 (that equals 6.91), where also As have the questionnaire task as source of earnings, is significantly higher than the average preference expressed in all other treatments. This results make sense since in Treatment 2 subjects are aware they have two sources of earnings (their performance in the questionnaire and the earnings from extortion) instead of one only (the earnings from extortion) as in the other treatments. Mann-Whitney rank-sum tests on significance differences are reported in Table 3 (the third column refers to preference for A role, whereas the first and the second to requests and compliance respectively. They will be commented on below.).

[TABLE 3]

Furthermore, we find a negative and significant correlation between subjects' preference for A's role and measures of their performance in the task: effort, as measured by the amount of seconds they save when answering to the questionnaire - that are converted into tokens in order to reward quick answers - and ability - that capture both their competence in answering correctly (Spearman correlation test, with coef. = -.118, p = .000, and coef.= -.317, p=.000).

Descriptive statistics on levels of requests (expressed as percentage of earnings) and compliance (expressed as percentage of the request) in the four treatments are summarized in Table 4.

[TABLE 4]

Overall, the level of the requests ranges between 31.9% and 48% of earnings, with major variations across treatments. Choosing the role of A instead of B determines a significant drop in A's average request. Interestingly, subjects who self-select into the role of extorter end up into making significantly lower requests: on average, the levels of percentage request are

35.6% in Treatment 1 (“Baseline”) and 31.9% in Treatment 2 (“Role choice”) (Wilcoxon rank-sum test on individual averages, with p = 0.065, two-tailed test; see Table 3). A possible rationale for this finding is the following: when choosing the role of A, individuals feel responsible towards their peers and make moderate requests; when randomly assigned to the role of A, they can shift the responsibility of acting greedily to external factors and find a justification for high requests. This result is consistent with the experimental evidence in Charness (2000), showing that a shift of responsibility to an external authority diminishes internal impulses towards prosocial behavior.

When the endowment is asymmetric, as in Treatment 3 (“Role choice + Asymmetric Endowment”), the average request significantly increases, rising to 45.7% (Wilcoxon rank-sum test on individual averages, with p = 0.000, two-tailed test; see Table 3): the explanation is related to equity concerns, since subjects A predate more because they have no other sources of earnings other than extortion.

Treatment 4 (“Feedback”) differs from Treatment 3 on subjects Bs’ set of information only: from the second round on, after each interaction B receives a feedback on the other B’s previous degree of compliance. Consistently with the fact that there is no difference in the set of actions and information for subjects As, the average request (48.0%) is not significantly different from the one in Treatment 3 (Wilcoxon rank-sum test on individual averages, with p = 0.541, two-tailed test).

Table 3 also shows that apparently there is no significant difference in the levels of compliance between treatments 1, 2, 3, and 4. However, compliance in the Treatment 4 (“Feedback”) contains an interesting dynamics. When subjects Bs are informed that in the previous period the other B in the group has offered more than 50% of requested tokens, they react by offering on average 56% of requested tokens. When they are informed that that in the previous period the other *e* in the group has offered less than 50% of requested tokens, they react by offering on average 27%. Both compliance levels (in percentages) are significantly different from the one observed in Treatment 3, that differs only for the fact that this information is not given to Bs. Victims’ compliance is therefore conditional to the behavior of other victims.

Table 5 provides a summary of the number of punishment points that subjects Bs receive in all the four treatments.

[TABLE 5]

In general, punishment is used in less than 30% of cases. We distinguish among low, medium and high intensity punishment, meaning that subjects Bs experience the elimination of one, two or three questions in the following round with the consequent impossibility to earn money by answering the question(s). There is no difference across Treatments 1, 2, 3 and 4 in frequency and intensity (high, medium and low) of punishment: overall, the number of punishment points decreases as interaction goes on, because fewer subjects inflict punishment (Wilcoxon rank-sum test on individual averages, with $p = 0.033$, one-tailed test). The use of punishment of low intensity is the most chosen strategy by extorters in all the treatments: it represents the 61.3% of punishment points whereas medium and high intensity punishment are used in 20.6% and 18% of cases respectively.

5.2. Regression analysis

The following regressions provide a deeper analysis of the determinants of A's request and of B's compliance, received punishment and performance in the task.

A's request

Table 6 shows that the level of the request significantly and positively depends on (1) the victim's earnings, (2) the request made to the other victim, (3) the presence of an asymmetric endowment (i.e. the extorter asks for more when extortion is the only source of income). The level of the requests depends significantly but negatively on (4) the other victim's earnings, (5) the period of interaction (the longer the interaction, the lower the request), (6) the choice of the role (when the subject can self-select into the role of extorter he turns to behave less greedily). We found neither gender effect, or any role for faculty and age.

[TABLE 6]

We therefore find evidence suggesting that A's appropriation request is proportional to subject B's earnings, since the coefficient is positive and significant. Furthermore, subject A makes similar requests to both Bs, showing a good degree of consistency in the behavior towards his potential victims and therefore exhibiting the same "extortive" style towards both of them.

Victims' compliance

Table 7 summarizes the determinants of the increase in percentage compliance across all the treatments.

[TABLE 7]

Subjects Bs' increase of compliance in time depends significantly and positively on (1) the A's request, (2) the victim's own earnings, and (3) the intensity of punishment received in the previous period. The regression confirms that treatments play no role. Subjects Bs increase compliance the higher their earnings and the higher the request made by A. Punishment appears to be effective in sustaining compliance. Notably, the level of compliance drops consistently in the last period, when punishment is not possible: the average level of compliance in the 10^{th} round is 17.5%.

Table 7bis and 8 provide a deeper investigation of the role of feedback in shaping compliance in Treatment 4.

[TABLE 7bis]

[TABLE 8]

When the feedback on the other victim's compliance is available, as happens in Treatment 4, each subject Bs shapes her own level on compliance on it. Interestingly, in presence of feedback punishment turns out to play no role (see Table 7bis), probably neglected because comparatively less salient.

When the feedback informs the B that the other B in the group has offered more than 50% of the request token, then the level of her compliance is higher than 50% and *viceversa* (see Table 8). While the difference in the level of percentage compliance between the two Bs is 42.03% in the first period, from the second period on it drops to 8% and remains stable across periods, with the exception of the last period, when it rises again and reaches 17.89%. Victims' compliance is conditional to the behavior of other victims.

Punishment

Table 9 investigates the determinants of the amount of punishment points that subjects Bs receive.

[TABLE 9]

Unsurprisingly, punishment depends significantly and negatively on (1) the previous degree of compliance: the more a subject B accomplished A's requests, the lower the number of punishment points he receives. Furthermore, punishment depends significantly and positively on (2) A's request, and significantly and negatively on (39 time: as the interaction goes on, punishment declines. Since the interaction between A and Bs assumes a form similar to a bargain that evolves in time, each couple A-B seems to converge to a level of compliance that satisfies both subjects and that requires less punishment: in fact, the number of punishment points received declines in time. As observed above, there is no difference in intensity of punishment across time and treatments.

Effects of extortion on victims' effort and ability

What is the effect of extortion and punishment of non-compliant victims on the overall economic activity of its victims? Is there any crowding-out effect? The following regressions show that the answer is no for what concerns extortion, but yes for what concerns punishment. In fact, in each period both effort and ability are significantly and positively affected by the amount of tokens extorted in the previous period, that also capture the victim's previous degree of compliance. However, the amount of punishment received in the previous period crowds out both effort and ability, as shown in Table 10 and Table 10bis.

[TABLE 10]

[TABLE 10bis]

5.3. Fair request and emotions

In the final questionnaire we ask subjects in both A and B roles to assess the level of request they judge as **"**fair**"**[4] and to express in a scale from 1 to 10 the intensity of a number of emotions they could experience during the experiment (happiness, guilt, shame, surprise, envy, anger, sadness, irritation, pride, gratitude, admiration, fear, blame, disappointment, grief, indignation)[5].

The average self-reported **"**fair request**"** is about 32% of earnings: it is worth noting that subjects As declare a percentage request that was on average 38%, whereas subjects Bs reported a percentage request of 28% (the difference is significant, Wilcoxon rank-sum test on

[4] See Binmore *et al.* (1991) for research on subjects' perception of fairness norms.

[5] The list of emotions and the methodology used for measuring them is based on Reuben and van Winden (2010).

individual averages, with p = 0.0001, two-tailed test). Interestingly, the fair request is significantly higher for subjects Bs in the Baseline, who show an average percentage of 40.7% compared to 25.7% of subjects Bs in the other treatments (Wilcoxon rank-sum test on individual averages, with p = 0.0001, two-tailed test): this could serve as confirmation that receiving the role of A at random and then behave consistently is judged less negatively in terms of responsibility also by subjects Bs. In fact, Bs tolerate a significantly higher request, and consider it as "fair" only in the treatment where As did not choose the role (i.e., the baseline).

The correlation analysis shows that for subjects Bs, compliance is significantly and positively correlated to the level of request judged as "fair" (coef. = 0.085, p=.0001). Furthermore, the distance between the average request received and the request judged as "fair" is significantly and positively correlated with irritation (coef. = 0.050, p=.039), and anger (coef. = 0.085, p=.0001).

For subjects As, happiness and gratitude increase when Bs' compliance increases (coef. = 0.262, p=.000; coef. = 0.152, p=.0001 respectively), whereas disappointment and blame decrease in Bs' compliance (coef. = -0.113, p=.001; coef. = -0.071, p=.037 respectively).

6. Discussion and conclusions

This paper presents an investigation of both the extorter's and the victim's behavior in the context of a new game, called "extortion game", that provides some important insights on the extent of how extortive requests are shaped by extorters, how punishment influences victims' compliance, and how victims react to extortive requests that are perceived as more or less fair. Furthermore, we explore the effects of informing the victims about the behavior of other victims.

Our results show that the extorter's requests are *proportional* to the victim's earnings, and are typically balanced to the victim's profit and similar across victims. This can be interpreted such as each extorter has his specific *extortive style* that is kept constant across victims. Furthermore, an extorter who self-selects into this role makes significantly lower requests showing some sense of responsibility and fairness towards their victim (with respect to an extorter who acquires the role randomly). Punishment is rare and the extorted sums are rather small, about 10% - 15% of the victim's earnings.

Results also shed light on the motivations that make a victim compliant with extortive requests: they confirm that punishment plays a role in promoting compliance, but they also show the importance of other concurrent factors.

Nevertheless, if punishment as the benefit obtained in return for the payment of extortion money was the only factor affecting victims' behaviors, then compliance with the extortive request could have been explained simply as an instrumentally rational act. However, our results show that victims are also sensitive to other factors that do not fit with this explanation. In particular, fairness perceptions play an important role: the requests made by extorters are positively correlated with the request they perceive as fair; victims who are required to pay what they consider to be an unfair amount show lower levels of compliance and experience higher intensities of emotions, such as anger and irritation, than when the request is perceived as fair. Moreover, information about other victims' compliance significantly affects compliance with extortive requests. Victims tune their compliance on the behavior of their peers, showing that conformity to others' conduct represents a strong and robust drive of human behavior (Carpenter, 2004; Moscovici, 1985; Cialdini and Trost, 1998) and that social norms influence individuals' conduct also in situations that are not beneficial for the society. Interestingly, these findings show that besides punishment, the perceived fairness of the request and others' compliance are key factors in eliciting individuals' obedience to the requests of an authority, even in contexts that are anti social or harmful. These may be considered as factors sufficient to induce obedience independently from the content of request and possibly from the specific nature of the authority.

References

Andrighetto, G., Brandts, J., Conte R, Sabater-Mir J, Solaz H, and Villatoro, D. (2013). Punish and Voice: Punishment Enhances Cooperation when Combined with Norm-Signalling. PLoS ONE, 8(6): e64941.

Bazart, C., & Bonein, A. (2014). Reciprocal relationships in tax compliance decisions. Journal of Economic Psychology, 40, 83–102.

Becker, G. (1968). Crime and Punishment: An Economic Approach. The Journal of Political Economy, 76, 169-217.

Bicchieri, C. (2006). The Grammar of Society. The Nature and Dynamics of Social Norms. Cambridge, MA: Cambridge University Press.

Binmore, K., Morgan, P., Shaked, A., and Sutton, J. (1991). Do people exploit their bargaining power? An experimental study. Games and Economic Behavior, 3, 295-322.

Camerer, C. (2003). Behavioral Game Theory. New Jersey: Princeton University Press.

Carpenter, J. (2004). When in Rome: Conformity and the provision of public goods. Journal of Socio-Economics, 33(4), 395-408.

Carter, J R., and Anderton, C. H. (2001). An experimental test of the predatory-prey model. Journal of Economic Behavior and Organization, 45, 83-97.

Charness, G. (2000). Responsibility and effort in an experimental labor market. Journal of Economic Behavior and Organizations, 42, 375-384.

Charness, G., and Haruvy, E. (2002). Altruism, equity and reciprocity in a gift exchange experiment: An encompassing approach. Games and Economic Behavior, 40, 203–231

Cialdini, R., Kallgren, C., & Reno, R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behaviour. In M. P. Zanna (Ed.), Advances in Experimental Social Psychology (Vol. 24, pp. 201–234). New York, NY: Academic Press.

Cialdini, R. and Trost, M. (1998). Social influence: Social norms, conformity, and compliance. In Gilbert, D., Fiske, S., and Lindzey, G. (Eds.). The Handbook of Social Psychology. Boston, McGraw-Hill. 2: 151-192.

Coleman, J. (1990). Foundations of Social Theory. Cambridge, MA: Belknap.

Conte, R., Andrighetto, G., and Campennì, M. (Eds.) (2014). Minding Norms. Mechanisms and Dynamics of Social Order in Agent Societies. New York: Oxford University Press.

Cox, J. C., Ostrom, E., Sadiraj, V., and Walker, J.M. (2013) Provision versus appropriation in symmetric and asymmetric social dilemmas. Southern Economic Journal, 79 (3), 496-512.

Ellsberg, D. (1968). The Theory and Practice of Blackmail. Rand, Santa Monica, CA.

Fehr, E. and Fischbacher, U. (2004). Social norms and human cooperation. Trends in Cognitive Sciences, 8(4), 185-190.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. Nature, 415(6868), 137–140.

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

Fehr E. and Schmidt, K.M. (2003). Theories of fairness and reciprocity - Evidence and economic applications. In: Published in: Dewatripont, M., Hansen L. P., and Turnovsky, S. J. (2003). Advances in Economics and Econometrics, Econometric Society Monographs, Eighth World Congress, Volume 1, 208-257.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a

public goods experiment. Economics Letters, 71(3), 397–404.

Gintis, H. (2000). Strong Reciprocity and Human Sociality. Journal of Theoretical Biology, 206, 169-179.

Greiner, B. (2004). An online recruitment system for economic experiments, MPRA Paper No. 13513.

Grossman, H.I., Kim, M. (1996). In Garfinkel, M., Skaperdas, S. (Eds.), Predation and Production: The Political. Economy of Conflict and Appropriation. Cambridge University Press, New York, 57-71.

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. Journal of Economic Behavior & Organization, 3, 367-388.

Moscovici, S. (1985). Social influence and conformity. In Gardner, L. and Aronson, E. (Eds.) Handbook of social psychology: Volume II Special fields and applications. New York: Random House. 347-412.

Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without a sword: Self-governance is

possible. The American Political Science Review, 86(2), 404–417.

Rueben, E., and Van Winden, F. (2010). Fairness perceptions and prosocial emotions in the power to take. Journal of Economic Psychology, 31, 908-922.

Tyler, T. (2006). Why People Obey the Law. Princeton: Princeton University Press.

Van Winden, F. (2001). Emotional hazard exemplified by taxation-induced anger. KYKLOS, 54, 491-506.

Walker, J., Gardner, R., and Ostrom, E. (1990). Rent dissipation in limited access common pool resource environments: experimental evidence. Journal of Environmental Economics and Management 19, 203-211.

Xiao, E., and Houser, D. (2011). Punish in public. Journal of Public Economics, 95, 1006-1017.

Figures and tables

Table 1: Treatments

	ROLE ASSIGMENT	ENDOWMENT	PUNISHMENT	FEEDBACK ABOUT THE OTHER B's COMPLIANCE
T1: BASELINE	Random	Symmetric	Yes	No
T2: ROLE CHOICE	Preference	Symmetric	Yes	No
T3: ASYMMETRIC ENDOWMENT	Preference	Asymmetric	Yes	No
T4: FEEDBACK	Preference	Asymmetric	Yes	Yes

Table 2: Average preference for A role

	1	**2**	**3**	**4**
	Baseline	**Role choice**	**Role choice + As End**	**Feedback**
Average preference for A role	-	6.91	5.04	5.15

Table 3: Significance levels of difference in requests, compliance and preference for A role

Fairness and Compliance in the Extortion Game

across treatments

Mann-Whitney Rank-sum test			
	Request	Compliance	Preference for A role
1 (Baseline) vs 2(Role choice)	**p=.065**	p=.180	-
2 (Role choice) vs 3 (Role choice+As End)	**p=.000**	p=.837	**p=.000**
3 (Role choice+As End) vs 4 (Feedback)	p=.541	p=.312	p=.687

Table 4: Descriptive statistics on percentage requests and compliance

	1	**2**	**3**	**4**
Variable (averages)	**Baseline**	**Role choice**	**Role choice + As End**	**Feedback**
Request level= Request/Earnings	35.6	31.9	45.7	48.0
Compliance level= Appropriation/ Request	28.6	34.3	33.1	32.9

Table 5. Punishment points received across treatments

# punishment points received	**1** **Baseline**	**2** **Role choice**	**3** **Role choice + As End**	**4** **Feedback**
0	190	162	196	201
1	68	56	57	57
2	15	15	25	25
3	27	7	19	17

Table 6. Determinants of A's request

Subject *j*'s request to subject *i*	**Coef.**	**(Robust**

		Std. Err.)
Subject *i*'s earnings	.415***	(.027)
Subject *j*'s earnings	-.250***	(.024)
Request to *j*	.673***	(.045)
Period	-.435*	(.245)
Role choice	-5.972*	(3.186)
Asymmetric endowment	7.683**	(3.224)
Feedback	1.370	(2.982)
Constant	-5.762	(4.674)
Observations	*570*	
R-sq.overall	*.719*	
Wald chi2	*688.23*	

*Note: GLS regression. Robust standard errors clustered at subject level are reported in parenthesis. *, ** and *** denote coefficients significantly different from zero at the 90% confidence, 95% confidence and 99% confidence. Controls include: gender, age and major. The xttobit regression (with lower limit equal to zero and upper limit equal to maximum request) reports almost identical results.*

Table 7. Determinants of B's increase in compliance in all treatments

Subject *i*'s % compliance in t - Subject *i*'s % compliance in t-1	**Coef.**	**(Robust Std. Err.)**
Punishment received by *i* in *t-1*	.058***	(.024)
Role choice	.001	(.015)
Asymmetric endowment	-.001	(.016)

Feedback	-.005	(.013)
Constant	-.081	(.141)
Observations	*811*	
R-sq .overall	*.015*	
Wald chi2	*9.94*	

*Note: GLS regression. Robust standard errors clustered at subject level are reported in parenthesis. Periods go from 2 to 10. *, ** and *** denote coefficients significantly different from zero at the 90% confidence, 95% confidence and 99% confidence. Controls include: gender, age and major. The xttobit regression (with lower limit equal to zero and upper limit equal to maximum level of compliance reports almost identical results.*

Table 7bis. Determinants of B's increase in compliance in Treatment 4

Subject *i*'s % compliance in t - Subject *i*'s % compliance in t-1	**Coef.**	**(Robust Std. Err.)**
Punishment received by *i* in *t-1*	.034	(.048)
Constant	.144	(.183)
Observations	*219*	
R-sq .overall	*.011*	
Wald chi2	*2.55*	

*Note: GLS regression. Robust standard errors clustered at subject level are reported in parenthesis. Periods go from 2 to 10. *, ** and *** denote coefficients significantly different from zero at the 90% confidence, 95% confidence and 99% confidence. Controls include: gender, age and major. The xttobit regression (with lower limit equal to zero and upper limit equal to maximum level of compliance reports almost identical results.*

Table 8. Determinants of B's compliance: role of feedback in Treatment 4

	Treatment 4		
Subject *i*'s compliance	**Coef.**	**(Rob. Std. Err.)**	**(Rob. Std. Err.)**
Subject *i*'s received request	.116*	(.080)	(.043)
Subject *i*'s earnings	.119***	(.041)	(.025)
Punishment received by *i* in *t-1*	.557	(.824)	(.659)
Information on j's compliance	*.186***	*(.066)*	*(.056)*
Period	-.069	(.428)	(.316)
Constant	9.499	(2.979)	(7.088)
Observations	*120*		
R-sq. overall	*.499*		
Wald chi2	*61.26*		

Table 9. Determinants of B's received punishment points

Punishment points received by subject *i in t*	**Coef.**	**(Robust Std. Err.)**
Request received by subject *i* in *t-1*	.018***	(.003)
Subject *i*'s compliance in *t-1*	-.064***	(.008)
Period	-.156***	(.026)
Role choice	-.332	(.415)
Asymmetric endowment	.352	(.416)
Feedback	-.158	(.390)
Constant	-.205	(1.485)
Observations	*1138*	
Log likelihood	*1138.76*	
Wald chi2	*87.94*	

*Note: xttobit regression. The dependent variable ranges from 0 to 3. Standard errors are reported in parenthesis. *, ** and *** denote coefficients significantly different from zero at the 90% confidence, 95% confidence and 99% confidence. Controls include: gender, age and major.*

Table 10. Determinants of B's effort in all treatments

Subject *i*'s effort	**Coef.**	**(Robust Std. Err.)**
Subject *i*'s compliance in *t-1*	.167***	(.039)
Punishment received by *i* in *t-1*	-1.970***	(.714)
Role choice	2.172	(2.251)
Asymmetric endowment	-13.023***	(1.749)
Feedback	.601	(1.592)
Constant	33.980	(10.022)
Observations	*1023*	
R-sq .overall	*.19*	
Wald chi2	*133.10*	

*Note: GLS regression. Robust standard errors clustered at subject level are reported in parenthesis. *, ** and *** denote coefficients significantly different from zero at the 90% confidence, 95% confidence and 99% confidence. Controls include: gender, age and major.*

Table 10bis. Determinants of B's ability in all treatments

Subject *i*'s ability	**Coef.**	**(Robust Std. Err.)**

Subject *i*'s compliance in *t-1*	.286***	(.082)
Punishment received by *i* in *t-1*	-7.858***	(2.129)
Role choice	4.478	(3.887)
Asymmetric endowment	-3.399	(3.622)
Feedback	3.586	(3.624)
Constant	73.883***	(20.606)
Observations	*1023*	
R-sq .overall	*.064*	
Wald chi2	*56.95*	

*Note: GLS regression. Robust standard errors clustered at subject level are reported in parenthesis. *, ** and *** denote coefficients significantly different from zero at the 90% confidence, 95% confidence and 99% confidence. Controls include: gender, age and major.*