



HAL
open science

Extortion in the laboratory

Friedel Bolle, Yves Breitmoser, Steffen Schlächter

► **To cite this version:**

Friedel Bolle, Yves Breitmoser, Steffen Schlächter. Extortion in the laboratory. *Journal of Economic Behavior and Organization*, 2011, 78 (3), pp.207. 10.1016/j.jebo.2011.01.005 . hal-00989521

HAL Id: hal-00989521

<https://hal.science/hal-00989521>

Submitted on 12 May 2014

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

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

Accepted Manuscript

Title: Extortion in the laboratory

Authors: Friedel Bolle, Yves Breitmoser, Steffen Schlächter

PII: S0167-2681(11)00019-9
DOI: doi:10.1016/j.jebo.2011.01.005
Reference: JEBO 2665



To appear in: *Journal of Economic Behavior & Organization*

Received date: 17-2-2010
Revised date: 20-12-2010
Accepted date: 12-1-2011

Please cite this article as: Bolle, F., Breitmoser, Y., Schlächter, S., Extortion in the laboratory, *Journal of Economic Behavior and Organization* (2008), doi:10.1016/j.jebo.2011.01.005

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Extortion in the laboratory

Friedel Bolle Yves Breitmoser* Steffen Schlächter
EUV Frankfurt (Oder) EUV Frankfurt (Oder) EUV Frankfurt (Oder)

December 20, 2010

Abstract

In a laboratory experiment, we study a finitely repeated game ($T = 15$) under complete information. In each round, P demands tribute (cash transfer) from A , A complies or refuses, and after refusals P may punish A . In equilibrium (payoff maximization), P does not punish and A refuses any positive demand. In the experiment, P punishes increasingly often and increasingly severely as she gains experience; most A s comply with P 's demands. The observations are compatible with linear spite. In a finite mixture model, the types of P and A in the subject pool are characterized. An A that is resistant to extortion (declines all demands) is very rare, and hence the threat of punishment in general is effective, but all A s either ignore actual punishment or react negatively to it. They accept to pay tribute but they are resistant to piecemeal expropriation.

JEL classification: C72, C91, D74

Keywords: laboratory experiment, mafia, extortion, punishment, finitely repeated game

*We are grateful to a co-editor and two anonymous referees for very helpful comments.

Corresponding author: Yves Breitmoser. email: breitmoser@euv-frankfurt-o.de. Postfach 1786, 15207 Frankfurt(Oder), Germany, Telephone/Fax: +3355534 2291/2390.

- laboratory experiment on finitely repeated extortion game
- P (rincipal) demands tribute from A (gent), and if A does not comply, P may punish A
- P punishes increasingly often and increasingly severely after rejections
- eventually, most A s comply with P 's demands
- finite mixture modeling of type heterogeneity is applied to understand behavior

Extortion in the laboratory

December 20, 2010

Abstract

In a laboratory experiment, we study a finitely repeated game ($T = 15$) under complete information. In each round, P demands tribute (cash transfer) from A , A complies or refuses, and after refusals P may punish A . In equilibrium (payoff maximization), P does not punish and A refuses any positive demand. In the experiment, P punishes increasingly often and increasingly severely as she gains experience; most A s comply with P 's demands. The observations are compatible with linear spite. In a finite mixture model, the types of P and A in the subject pool are characterized. An A that is resistant to extortion (declines all demands) is very rare, and hence the threat of punishment in general is effective, but all A s either ignore actual punishment or react negatively to it. They accept to pay tribute but they are resistant to piecemeal expropriation.

JEL classification: C72, C91, D74

Keywords: laboratory experiment, mafia, extortion, punishment, finitely repeated game

1 Introduction

Many economic interactions are asymmetric in that one player, P (principal), is able to punish misbehavior of the other, A (agent), but effectively not vice versa. Such interactions can be found in social, military, and business hierarchies, between teachers and students, parents and children, or the mafia and shopkeepers. Formal analyses of such interactions usually build on the pair of assumptions that (i) punishment is sufficiently severe to destroy the benefits of misbehavior of opponents, and (ii) it is costly. The former implies that misbehavior is ruled out if the threat of punishment is credible, while the latter implies that it is non-credible in finitely repeated interactions. In turn, if the threat of punishment is non-credible (e.g. if the interaction has finite time horizon), A has nothing to fear and misbehaves: students harass their teachers, children mess around, shopkeepers refuse payment, and so on. Experimental analyses of this prediction (under complete information) are rare, however. The only dedicated experiment seems to be the one of Jung et al. (1994), who consider a game of P against a sequence of short-lived A 's, but in the above examples, A is actually long-lived.¹

In our experiment, P faces a long-lived A . The subjects are matched to form pairs and play a finitely repeated extensive form game. The time horizon ($T = 15$) is common knowledge. In each round, P is endowed with 40 and A is endowed with 160. They play a constituent game with three stages. First, P chooses a demand $x \in [0, 160]$, second A decides whether to accept or decline, and third, if A declined, P may punish A through reducing his income by some amount $y \in [0, 160]$. Punishment ($y > 0$) bears costs of 10 for P . The game has been designed to distinguish demands that exploit P 's strategic advantages from demands inspired by fairness concerns, and similarly to distinguish punishment in general from punishment to level incomes. For example, if P intends to level incomes, then she demands $x = 60$ ($\neq 160/2$, these choices are separated to distinguish whether P wishes half of A 's endowment rather than equality). Demands $x > 60$ suggest that P intends to exploit the strategic asymmetry and demands $x < 60$ suggest that P feels weak. Analogously, P punishes in order to level incomes

¹There is more loosely related work, e.g. on entry deterrence without the possibility of punishment ex post (Mason and Nowell, 1998), chain store games under incomplete information (Cooper et al., 1997), repeated "reputation" games under incomplete information, e.g. Camerer and Weigelt (1988), Neral and Ochs (1992), and Andreoni and Miller (1993), and also the wealth of experiments on ultimatum games, but in none of these cases, P may punish A in every round and the threat of doing so is non-credible.

by choosing $y = 130$ in stage 3, she punishes “adequately” in relation to the rejected demand by choosing $y = x$ (i.e. thus A does not profit through rejecting the demand), while the maximal and perhaps most effective punishment is $y = 160$.

The results can be summarized as follows. First, the principals (P) become increasingly inclined to punish ($y > 0$) as they gain experience, and eventually, they almost universally choose the maximal punishment $y = 160$ after declines of their demands. This inclination to punish is a divergence from subgame perfection² that almost all subjects in the role of P adopt. There is no “end effect” in that behavior approaches subgame perfection in the last rounds of the repeated game. Second, the principals’ demands are fairly moderate, often just high enough to equalize payoffs, and even such demands get rejected by some of the agents. This contrasts strikingly with the observed convergence toward maximal punishment. For, taking the latter as given essentially transforms the interaction into a repeated ultimatum game. In relation to ultimatum games, equalizing demands are moderate indeed and show that the P s struggle to fully exploit their strategic advantage.

Third, we observe that both demands and accepted demands are significantly increasing in the course of time, while the econometric estimates of the demand strategies are independent of “exogenous inputs” such as time, acceptance decisions, and previous punishment. Econometrically, they simply constitute stationary ARMA processes. To understand the discrepancy between overall trend and individual behavior, we re-estimate the strategies controlling for subject heterogeneity (in a finite mixture model). After segregating subject types, we find that that about 43% of the P s actually do increase demands in time, and another 42% of them increase demands after acceptance (while acceptance rates happen to be increasing in time). In turn, only controlling for subject heterogeneity disentangles the diversity of behavior and explains the overall trend. Arguably, similar diversity may be hidden in many experimental analyses.

Fourth, our results confirm the suspicion that punishment is a two-edged sword (see e.g. Fehr and Rockenbach, 2003). The anticipation of getting punished induces agents to accept moderate demands, but the actual execution of punishment does not improve acceptance rates. About 40% of the agents ignore the level of punishment, and the remaining 60% react negatively to getting punished (i.e. their acceptance rates

²Unless stated otherwise, by “subgame perfection” we refer to the subgame perfect equilibrium (SPE) assuming that it is common knowledge that all players maximize payoffs.

drop). Overall, agents seem willing to comply with moderate requests, and in particular they give more than they would without the threat of punishment (i.e. in dictator games). But they ignore punishment as such, or at least seek to maintain a reputation of ignoring it, seemingly to prevent piecemeal expropriation. This resistance to punishment is sustained although P may punish at low costs, which suggests that it would be stable also under other cost structures.

Finally, let us comment on the persistence of punishment, despite its violation of subgame perfection.³ This has been observed similarly in the context of “altruistic” punishment by Fehr and Gächter (2000, 2002). Altruistic punishment differs from “egoistic” punishment in extortion games, however. Altruistic punishers intend to sustain efficiency in contributions to public goods, while egoistic ones intend to enforce redistribution of resources (or contracts, see e.g. Fehr et al., 1997). This affects the strategic consequences fundamentally, as for example Masclet et al. (2010) report that threats of altruistic punishment effectively increase contributions, while in our case the previous period’s punishment (= the threat for punishment in this period) decreases acceptance rates. The reason for the persistence of punishment in our experiment is therefore understood best in contrast to Jung et al. (1994), where punishment also serves redistribution but disappears over time. The reason for the divergence seems to be the relative cost-effectiveness. In our case, a self-chosen damage can be inflicted at costs of 10, while Jung et al. give P the option to destroy 70 at costs of 90 for herself. The low costs of punishment in our experiment loosely reflect the examples cited above (e.g. mafia or parental punishment), but analyzing experiments with intermediately effective technologies may allow future research to locate the threshold between persistence and disappearance of punishment. As for altruistic punishment, Egas and Riedl (2008) and in particular Nikiforakis and Normann (2008) have shown that its cost effectiveness is decisive.

The remainder of the paper is organized as follows. Section 2 describes the experimental game and the procedure. Section 3 describes the basic results. Section 4 estimates the subjects’ strategies and the subject types. Section 5 discusses the results and concludes.

³In most cases, convergence toward Nash equilibrium is observed in the final rounds of finitely repeated games, e.g. in Prisoner’s dilemmas (Selten and Stoecker, 1986), public goods games (Muller et al., 2008; Neugebauer et al., 2009), market games (Loomes et al., 2003), and investment games (Cochard et al., 2004).

2 The experiment

2.1 The game and equilibrium predictions

The experimental game is a finitely repeated game ($T = 15$) of an extensive-form constituent game. The players are matched with the same partner for the duration of the experiment and the role assignments (principal and agent) are held constant. The number of rounds ($T = 15$) had been chosen such that subjects can be considered experienced in the final rounds, without risking that the subjects perceive it as an infinitely repeated game at any point. Thus, behavior is projected to stabilize in the sense of Nash equilibrium in the final rounds, which allows us to assess individual strategies and implicitly the possible interdependence of preferences.

The constituent game consists of three decision stages.

0. The principal P is endowed with 40 Euro-cent, the agent A is endowed with 160 Euro-cent
1. P chooses a demand $x \in [0, 160]$
2. A decides whether to accept or reject the demand; if A accepts, then the round ends and the payoffs are $\pi_P = 40 + x$ and $\pi_A = 160 - x$
3. If A rejects, then P chooses a punishment $y \in [0, 160]$; if P chooses $y = 0$, then $\pi_P = 40$ and $\pi_A = 160$ results, and otherwise $\pi_P = 30$ and $\pi_A = 160 - y$ results

We refer to this game as the *extortion game*, but alternative interpretations are possible.

The equilibrium predictions depend on the players' utilities. If players maximize pecuniary payoffs (i.e. if $u_i = \pi_i$ for $i = A, P$) and act in accordance with subgame perfection, then the equilibrium payoffs are unique: $\pi_P = 40$ and $\pi_A = 160$. Along the equilibrium path, P demands any $x \in [0, 160]$, A rejects if $x > 0$, and P chooses zero punishment $y = 0$. This equilibrium prediction holds invariantly in all repeated games with finite time horizon.

Next, consider Fehr-Schmidt inequity aversion (Fehr and Schmidt, 1999). If $i \neq j \in \{P, A\}$ denotes the players and (π_i, π_j) the payoffs, then the Fehr-Schmidt utility of i is (using the guilt weight $a_i \geq 0$ and the envy weight $b_i \geq a_i$)

$$u_i = \pi_i - a_i \cdot \max\{\pi_i - \pi_j, 0\} - b_i \cdot \max\{\pi_j - \pi_i, 0\}. \quad (1)$$

An inequity averse P with Fehr-Schmidt preferences punishes non-compliance of A in equilibrium if and only if $b_P \geq 1/12$. If $b_P > 1/12$, then P chooses $y = 130$ in every SPE, and $(30, 30)$ results when stage 3 is reached. In Stage 2, A will therefore accept certain positive demands. For example, an inequity averse A knowing $b_P > 1/12$ generally accepts 60, since he prefers $(100, 100)$ to $(30, 30)$, but he would never accept a demand greater than 130. The demand x that makes A indifferent between accepting and rejecting is the maximal demand that he would accept.

$$x_{\max} = (130 + b_A \cdot 120)/(1 + 2b_A) \quad (2)$$

Hence, x_{\max} also is the choice of P in all SPEs if b_A is common knowledge (assuming $a_P \leq 1/2$; she demands $x = 60$ if $a_P > 1/2$). It equates with 130 if A is egoistic ($b_A = 0$) and it approaches 60 as b_A tends to infinity.⁴

Finally, assume that the players' preferences exhibit "linear altruism" $u_i = \pi_i + \alpha_i \pi_j$ for $i \in \{A, P\}$. In this case, P punishes a non-compliant agent if $\alpha_P < -1/16$, and if she does, she chooses maximal punishment $y = 160$. Anticipating such punishment, in the unique SPE, an A with $\alpha_A \geq 1$ would accept any demand, and an A with $\alpha_A < 1$ would accept demands up to

$$x'_{\max} = (160 + 10\alpha_A)/(1 - \alpha_A). \quad (3)$$

That is, an altruistic agent ($\alpha_A \geq 0$) would accept any demand, and spiteful agents with for example $\alpha_A = -10/7$ would accept demands up to 60. In contrast to inequity aversion, a spiteful principal does not generally prefer the equitable payoff allocation $(100, 100)$ over the disagreement payoff, which is $(30, 0)$ if P is spiteful, nor does she prefer the maximal sustainable allocation $(160 - x'_{\max}, 40 + x'_{\max})$ over disagreement in general. If she does not, namely if $x'_{\max} < (-10 - 160\alpha_P)/(1 - \alpha_P)$, then P deliberately makes an offer that A will reject.

There are at least two other reasons why P may punish non-compliance of A : P may try to build a reputation in the sense of Kreps and Wilson (1982) and Milgrom and Roberts (1982), or P may be boundedly rational. The former may be relevant even in a game that is repeated only 15 times, e.g. if the share of boundedly rational players to be mimicked is sufficiently high, which seems reasonable in laboratory experiments.

⁴If b_A is private information of A , then P maximizes her expected utility according to her risk attitude and the distribution of b_A in the population, and ends up choosing some $x \in [60, 130]$ as well.

2.2 Experimental procedure

The experiment took place in the computer laboratory of the European University Viadrina in Frankfurt (Oder), Germany. Each computer terminal in the laboratory was partitioned, so that subjects were not able to look at other computer screens, or to communicate via audio or visual signals. Subjects were recruited from an email list consisting of students from the faculties of Cultural Science, Business and Economics, and Law. We conducted eight sessions, each with either 10 or 12 participants. Altogether, 92 subjects were randomly and anonymously matched to form 46 pairs who played 15 rounds of the above game. Our data set therefore comprises observations from 46 stochastically independent finitely repeated games.

Each session proceeded as follows. First, subjects were randomly allocated to their seats, second they were provided with the experimental instructions and a short control questionnaire (translations of both are provided in the appendix). Their answers to the control questionnaire allowed us to verify their understanding. Subjects in doubt were verbally advised by assistants before the experimental game began. At the end of the experiment, subjects were informed of their payments, and asked to privately choose a code name and password. This was used to anonymously collect their payments from a third party about a week after the experiment.

Throughout the experiment, “mildly” loaded language was used. For example, the game was not referred to as “mafia game” or “extortion game” and the term “punishment” was avoided, but we describe the stage-3 action of P as “If the A -participant rejects the demand, then the B -participant has the option to destroy an arbitrary number of points (maximally 160) of the A -participant.” The experimental instructions and the full data set are provided as supplementary material. The experiment was conducted using z-tree (Fischbacher, 2007), the programs are available from the authors upon request. Every session lasted about 30–40 minutes. The average income was € 12.07 and varied between € 6.45 and € 18.45.

3 Basic results

The average results are displayed in Figure 1. Almost all demands are positive (99%), and the most frequent demand is 60. When accepted, a demand of 60 equates the final

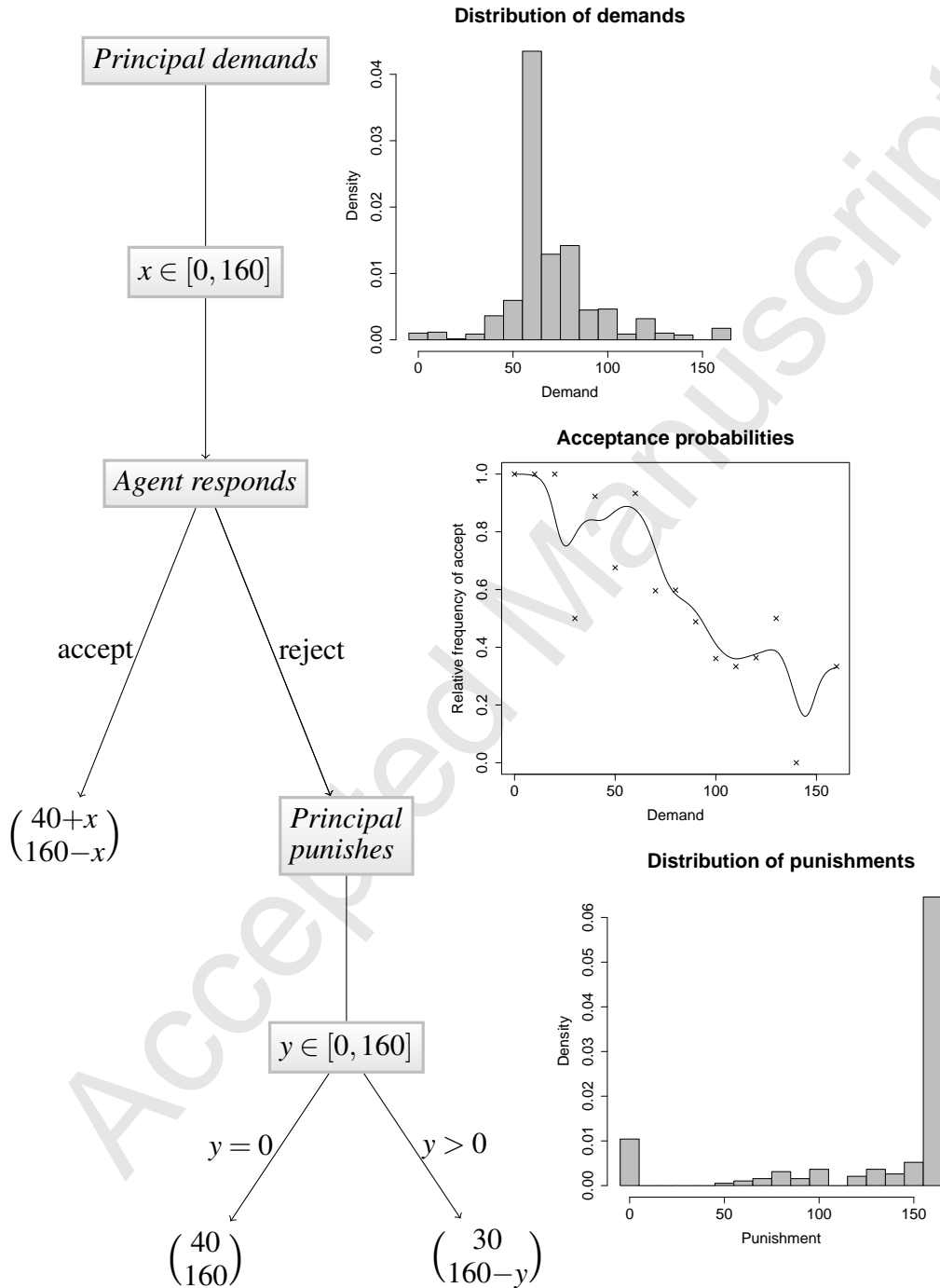
incomes at 100 each. Principals usually punish rejections of their demands (in 90% of the cases), and in 70% of these cases the maximal punishment $y = 160$ had been chosen. The occurrences of maximal punishment are compatible with linear altruism if $\alpha_P < -1/16$. Punishment below the principal's demand ($0 < y < x$) was observed in only 3.6% of the cases, and similarly $y = x$ (relative frequency 1%) or the SPE punishments of egoistic players ($y = 0$, relative frequency: 10.4%) and Fehr-Schmidt players ($y = 130$, 3.6%) had not been chosen systematically. Apparently anticipating severe punishment, agents typically accept moderate demands. Overall, 72% of the demands were accepted, while the average accepted demand was 64.64, and the average rejected demand was 85.38. These values differ significantly ($p < .001$),⁵ and in this sense agents are more likely to reject high demands. As we will see below, demands and acceptance decisions are compatible with linear altruism, too.

The paths of play of the 46 pairs are highly diverse. In some cases, the agent seems to try to enforce demand reductions by rejecting previously accepted demands, in other cases, the principal seems to be attacking (by raising demands after acceptances), and in some pairs, the demands had actually been held constant for all 15 rounds and got accepted throughout. Such variations in the course of action, and its structure, suggest that the subject pool is not homogenous. This will be discussed in detail once the basic structure of the data over time has been described.

Figure 2 presents box-and-whisker plots of demands and accepted demands over time. It shows that the quartiles of the demands remained fairly constant for the duration of the experiment, and in particular the lower quartile is at $x = 60$ from round 3 on. Overall, demands are increasing in time, however. Tables 1–3 describe how demands, acceptance rates, and punishment changed over time. In these three tables, we have segregated the 15 rounds into three phases of five rounds each. Table 1 shows that the relative frequency of demands $x < 60$ decreases over time, from 23% initially to 6% eventually, and the relative frequency of demands $x > 60$ increases over time, from 37% initially to 55% eventually. These changes are significant ($p < .001$ and $p < .001$, respectively, in Wilcoxon matched pairs tests, using the 46 pairs as independent observations). In parallel to this development, the “accepted demands” changed over time. Figure 2 shows that the eventual lower quartile is above the initial upper quartile. Similarly, Table 2 shows that the acceptance rates with respect to most levels

⁵in a Mann-Whitney- U test using the mean demands within the 46 pairs as independent observations

Figure 1: Game time line and average decisions



Note: The second graph (“Acceptance probabilities”) displays the relative frequencies of acceptance for all intervals $[0, 5), [5, 15), \dots, [145, 155), [155, 160]$, and in addition it shows the Nadaraya-Watson kernel regression estimate of the acceptance function with “normal” kernel and bandwidth 20.

of demands increased over time.

Overall, the principals' incomes increase from 79.6 initially to 90.1 eventually ($p < .01$ in a Wilcoxon paired sample test, using the $n = 46$ individual means as independent observations) and the agents' incomes decrease from 84.1 to 71.2 ($p = 0.046$), see Table 1. This shows that the extortion games converge to a state favoring P , but the sources of this development (i.e. the time dependence of demands and acceptance rates) are not obvious. This is analyzed in the next section.

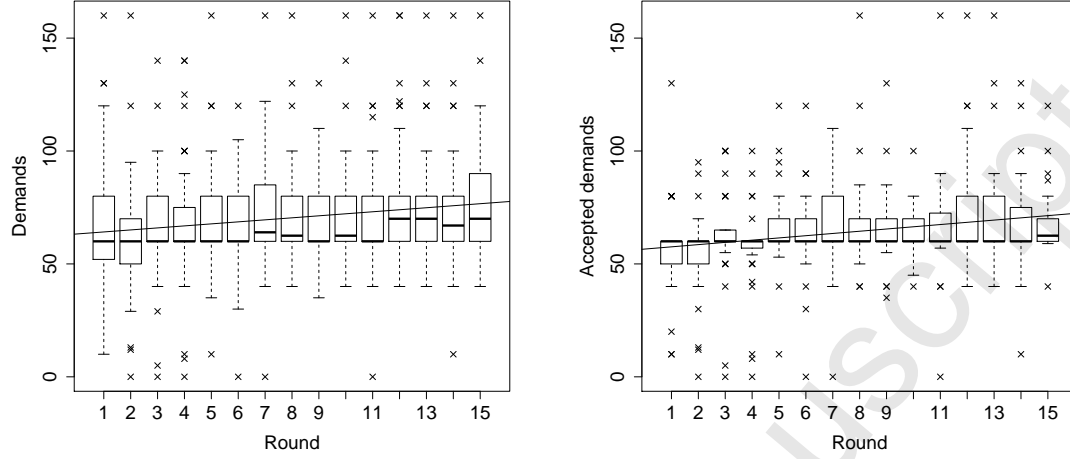
The aggregate behavior does not converge toward the exact predictions for either payoff maximizing players or Fehr-Schmidt players. The former is incompatible with the observation that both demands and acceptance rates are clearly positive and even increasing in time, and the latter is not exactly compatible with the joint distribution of demands and punishments. In the last third of the experiment, most subjects choose the maximal punishment $y = 160$ (82%, see Table 3)—when punishment does not serve reputation building anymore—while the majority (55%) of demands is $x > 60$. Demands $x > 60$ suggest that most players are at most partially inequity averse, but if so, they should punish by $y \leq 130$. A possible explanation is that principals are inequity averse in general and spiteful after rejections (i.e. angered), but our experiment was not designed to distinguish this from the simpler notion that P s are spiteful throughout. Specifically, maximal punishment results for $\alpha_P < -1/16$ in linear altruism, demands $x \geq 60$ result for $\alpha_P \geq -7/10$, and the intersection of these conditions is non-empty.

4 Strategy estimation and latent type classification

To understand why extortion is profitable and how individuals act in detail, we now estimate the subjects' strategies. For a wide range of utility functions, the SPE predictions are unique for finitely repeated extortion games. This follows from the uniqueness of the constituent game equilibria, and implies in our case that the SPE strategies are even stationary (i.e. time invariant). Assuming stationarity seems a little restrictive when estimating behavioral strategies, however, and we therefore allow for a variety of alternative influences.

In modeling the demand strategy $(X_{i,t})_{t=1}^{15}$ of principal i , we allow for correlation with last round's demand $X_{i,t-1}$ (and for this reason, we skip the observation for $t = 1$),

Figure 2: Box-and-whisker plots of demands and accepted demands



Note: The box covers the interquartile range, and the whiskers extend to the most extreme data points that are within 1.5 times the interquartile range of this box (default in GNU R).

Table 1: Demand categories and incomes over time

Rounds	Rel. frequency of demand x			Incomes		
	$x < 60$	$x = 60$	$x > 60$	Agent	Principal	Σ
1–5	22.6%	40.0%	37.4%	84.1	79.5	163.7
6–10	12.6%	38.3%	49.1%	75.0	82.8	157.8
11–15	6.1%	38.7%	55.2%	71.2	90.2	161.3

Table 2: Frequencies of demands and acceptance rates (in brackets) over time

Rounds	$[0, 45)$	$[45, 55)$	$[55, 65)$	$[65, 75)$	$[75, 85)$	$[85, 105)$	$[105, 160]$
1–5	26(89%)	23(74%)	95(86%)	24(58%)	25(56%)	22(50%)	12(17%)
6–10	12(92%)	13(62%)	93(97%)	36(50%)	31(55%)	31(39%)	11(36%)
11–15	8(100%)	1(0%)	96(97%)	34(71%)	41(66%)	26(42%)	18(44%)

Table 3: Frequencies of punishment (conditional on demand rejection)

Rounds	$y = 0$	$y \in (0, 160)$	$y = 160$	Σ
1–5	10 (14.9%)	26 (38.9%)	31 (46.3%)	67
6–10	7 (10.1%)	18 (26.1%)	44 (63.8%)	69
11–15	3 (5.4%)	7 (12.5%)	46 (82.1%)	56

with last round's acceptance decision $A_{j,t-1} \in \{0, 1\}$ of her agent j , with the number of rounds left $R_t = 15 - t$, with last round's punishment $P_{i,t-1}$, and for serial correlation of individual errors $\varepsilon_{i,t-1}$. Such econometric models are known as censored "ARMAX" models (auto-regressive moving average with exogenous inputs); the censoring follows from the bounds of the strategy set. The estimated coefficients and intercept, including their standard errors are as follows.⁶

$$X_{i,t} = \underset{(3.9259)}{3.017} + \underset{(0.0314)}{0.9289^{***}} \cdot X_{i,t-1} + \underset{(3.039)}{4.1128} \cdot A_{j,t-1} - \underset{(0.1021)}{0.0578} \cdot R_t \\ + \underset{(0.021)}{0.0084} \cdot P_{i,t-1} - \underset{(0.0532)}{0.5584^{***}} \cdot \varepsilon_{i,t-1} + \varepsilon_{i,t}$$

where $\varepsilon_{i,t}$ has standard deviation $\hat{\sigma}_\varepsilon = 18.87$. This suggests that demands are best described as ARMA processes—the "exogenous inputs" such as acceptance decisions, time, and punishment are insignificant with respect to the "representative" subject. The obvious issue with this conclusion is that the assumed existence of a "representative" subject is invalid. For this reason, we now relax the assumption of subject homogeneity. The econometric technique to do so is finite mixture modeling (Peel and MacLahlan, 2000). In the existing economic literature, finite mixture modeling is still used irregularly, but the applications that exist have shown that latent heterogeneity is a concern in experimental analyses. For example, Bardsley and Moffatt (2007) and Fischbacher and Gächter (2010) show heterogeneity of strategies in public goods contributions, Conte et al. (2008) and Harrison and Rutström (2009) show heterogeneity of preferences in choice under risk, and heterogeneity of "levels of reasoning" in p -beauty contest is observed by Stahl (1996) and Ho et al. (1998) and in other games by Kübler and Weizsäcker (2004) and Crawford and Iriberry (2007), to name just a few. These studies suggest that heterogeneity of subject pools is the rule rather than the exception, and thus also that finite mixture modeling should be considered routinely.

Technically, we consider a finite mixture of tobit models to represent individual demand strategies (similar to e.g. Bardsley and Moffatt, 2007, for public goods contributions) and finite mixtures of logit models to represent acceptance strategies. The details are summarized in Appendix A. The results are in Tables 4 and 5.

Before we discuss the estimated demand strategies, let us note that the Bayes information criterion (BIC, Schwarz, 1978) improves by about 200 points per subject

⁶In this section, we use "*" to denote significance at $\alpha = .05$ (in two-sided tests), "***" to denote significance at $\alpha = .01$, and "****" to denote significance at $\alpha = .001$.

Table 4: Maximum likelihood estimates of the proposer models

Model	Type	Interc	Dem	Accept	RoundsLeft	Punish	Theta	Sigma	Share	BIC _{LL}
A	1	3.017 (3.9259)	0.9289*** (0.0314)	4.1128 (3.039)	-0.0578 (0.1021)	0.0084 (0.021)	-0.5584*** (0.0532)	18.8722*** (0.5258)		2828.33 (-2805.69)
B	1	9.9768 (7.5578)	0.9033*** (0.0574)	2.3947 (5.2418)	-0.2064 (0.2181)	-0.0305 (0.0357)	-0.5754*** (0.092)	26.0469*** (1.1605)	0.4667*** (0.0766)	2638.34 (-2589.83)
	2	-3.4758 (3.1001)	0.9389*** (0.0291)	8.6467*** (2.3766)	-0.0573 (0.0705)	0.0549*** (0.0164)	-0.3431*** (0.0576)	6.9881*** (0.3628)		
C	1	55.7812** (17.7765)	0.3865 (0.2156)	0 (-)	-0.8942 (0.5668)	0 (-)	-0.0433 (0.2282)	32.4928*** (1.7722)	0.2865*** (0.069)	2448.12 (-2373.74)
	2	13.6486*** (0.0032)	0.7726*** (2e-04)	0 (-)	0 (-)	0 (-)	-0.7119** (0.2173)	0.1724*** (0.0151)	0.108* (0.0456)	
	3	9.0971* (4.1901)	0.7426*** (0.0547)	10.6469*** (2.8261)	0 (-)	0.0513** (0.0193)	-0.0354 (0.0864)	10.6122*** (0.4943)		

Table 5: Maximum likelihood estimates of the responder models

Model	Type	Int	Dem	ΔDem	RoundsLeft	Punish	Share	BIC _{LL}
A		4.2034*** (0.4319)	-0.03187*** (0.00483)	-0.02716*** (0.00723)	-0.07437** (0.02311)	-0.00857*** (0.00149)		348.72 (-332.38)
B	1	7.8422*** (1.4654)	-0.0505*** (0.0104)	0.0115 (0.0084)	-0.2435** (0.0865)	0.0046 (0.0059)	0.2958*** (0.0894)	342.66 (-306.71)
	2	5.3390*** (0.6966)	-0.0520*** (0.0087)	-0.0621*** (0.0132)	-0.0677* (0.0299)	-0.0097*** (0.0019)		
C	1	7.6883*** (1.4034)	-0.0452*** (0.009)	0 (-)	-0.2736*** (0.0828)	0 (-)	0.2874*** (0.0825)	337.05 (-294.56)
	2	4.0678* (1.8914)	-0.0425 (0.0256)	-0.1132* (0.0491)	-0.2697* (0.1088)	0 (-)	0.0974 (0.0557)	
	3	6.9618*** (0.9611)	-0.078*** (0.0131)	-0.0501*** (0.0143)	0 (-)	-0.0073** (0.0024)		

type that we add (until level 3). Considering that improvements of about 20 points are highly significant in adding six-parametric types (according to likelihood ratio tests), improvements of 200 points are striking and confirm the observations on heterogeneity made in the aforementioned studies. In particular, note the significance of heterogeneity in contrast to the insignificance of exogenous inputs such as $A_{j,t-1}$, R_t , and $P_{i,t-1}$.

We identified three types of principals in our subject pool (according to the BIC). After eliminating insignificant parameters (again according to BIC) from the type definitions, we obtained model *C* reported in Table 4. Type 1 has relative frequency 28.7% and a demand strategy with residual standard error $\hat{\sigma}_{\epsilon_1} = 32.49$. That is, about two in seven subjects are of type 1. The long-term mean of $X_{i,t}^1$ is $\mu^1 = 56.8399 / (1 - .3658) = 90.9$, which is high in relation to the other two types. Hence, we refer to them as *aggressive* subjects. About one in nine subjects (10.8%) is of type 2— with the extremely low residual standard error $\sigma_{\epsilon_2} = 0.172$. Due to the low residual variance, and the long-term mean at $\mu^2 = 60$, these subjects will be referred to as *fair* subjects. The remaining approximately three in five subjects (60.6%) are of type 3. They have the residual standard error $\sigma_{\epsilon_3} = 10.61$, and most interestingly they strategically react to acceptance decisions (raising the demand by about 10 after acceptance). The long-term mean in response to rejecting agents, i.e. if $A_{j,t-1} = 0$ and $P_{i,t-1} = 160$, is $\mu^3 = 67.2$. The long-term mean in response to accepting agents, i.e. if $A_{j,t-1} = 1$ and $P_{i,t-1} = 0$, is $\mu^3 = 76.7$. We refer to them as *adaptive* subjects.⁷

The agent strategies are estimated similarly. Agent j responds with either “accept” or “reject” to the principals demand, which we represent as a finite mixture of logit models with the following independent variables: the current demand $X_{i,t}$, the demand change $\Delta_{i,t} = X_{i,t} - X_{i,t-1}$ in relation to the previous round, the number of rounds left $R_t = 15 - t$, and last round’s punishment $P_{i,t-1}$. Again three types are identified.⁸ The estimates are reported in Table 5, and the average marginal effects required to interpret the estimated strategies are reported in Table 6. The average ac-

⁷Since only 3/5 of the subjects react to rejections and punishments, this explains why these terms had been insignificant in the global regression model.

⁸A referee raised the concern that the behavior (and type) of a given A depends on behavior (type) of the P he is matched with, or vice versa. After computing the posterior classifications for all subjects, see Appendix A, we verified the stochastic independence between A ’s classification and P ’s classification in linear probability regression models. We found that there was no significant interaction effect between the types (at the 10% level), i.e. the classifications of A and P do not violate stochastic independence.

Table 6: Average marginal effects in the logit models for A's acceptance

Type	Demand $X_{i,t}$	Change $\Delta_{i,t}$	Rounds left R_t	Punishment $P_{i,t-1}$
1	-0.00378	0	-0.02291	0
2	-0.00637	-0.01698	-0.04043	0
3	-0.00919	-0.00590	0	-0.00086

Note: The marginal effect is the first derivative of the acceptance probability with respect to the respective independent variable. We computed these derivatives for all data points and report their averages.

ceptance probabilities of types 1 through 3 are 0.873, 0.388, and 0.711 respectively (these probabilities are the predictions averaged over all data points). Overall, Type 1 (share 28.8%) is largely *acceptive* with little response to the demand itself and intermediate response to the number of rounds left. Type 2 (share 9.75%) is on average *rejective* with intermediate (yet insignificant) response to the demand, and stronger response to demand changes and the number of rounds left. Finally, the majority of agents is of type 3 (share 61.5%). They are particularly *responsive*, with intermediate acceptance rates and strong negative responses to the level of the demand, to demand increases, and to punishment. The behavior of type 3 can be interpreted as reciprocal and the negative response to punishment resembles counter-punishment in the sense of Nikiforakis (2008). Note that the counter-punishment is substantial. After maximal punishment $y = 160$, acceptance rates drop by roughly 14% on average.

These strategy estimates allow us to shed light on the observations made in Section 3. Table 2 showed that for most levels of demands, the acceptance rates are increasing in time. This increase of acceptance probabilities in time has been found for the agents classified as “acceptive” or “rejective”. The majority of agents is classified as “responsive,” however, and they respond to the demand, to demand changes, and to punishment, but not to time. The univariate tests also had shown that principals' demands are increasing in time, while the estimated representative demand strategy was stationary. After disaggregating subject types, we can now see that time is relevant only for aggressive principals. In addition, “adaptive” principals increase demands (by about 10 points) when last round's demand had been accepted, and since acceptance rates are increasing in time, this implies that “adaptive” demands correlate with time, too—without causality. Finally, principals classified as “adaptive” seem to believe that

punishment improves acceptance rates (see that $P_{i,t-1}$ is significant for them). Agents do not respond positively to punishment, though. The first two types of agents do not respond to punishment at all, and the majority of agents (the “responsive” ones) react negatively.

5 Conclusion

In our experiment, a powerful player called *P*(rincipal) interacts with a second player called *A*(gent) for 15 rounds. In each round, *P* may demand a transfer from *A*, e.g. in terms of money, housework, homework, quality time, and so on. *A* may accept or decline the transfer, and after declines, *P* may punish *A* at comparably small costs for herself. The induced valuations have it such that subjects maximizing pecuniary payoffs do not make any transfer from *A* to *P* under subgame perfection, and that *P* would never punish *A* after rejections (due to the costs). We found that *P*'s demands are largely moderate, *As* tend to accept moderate demands, but *As* that decline a demand are punished harshly. Furthermore, one type of *Ps* (accounting for 3/5 of the population) raise demands after harsh punishment, seemingly expecting *A* to give in, and yet all types of *A* respond non-positively to punishment. The acceptance rates remain either constant or drop after punishment.

Overall, however, the subjects' interactions do not display excessive disagreement. The maximal combined earnings of *P* and *A* are € 30 in the experiment, and the average combined earnings were € 24.14—about 80% of the maximum, disagreement occurred on average only every fourth round. Hence, the fairness standards of the various *P* and *A* in our experiment are largely compatible with one another, which one would not expect unless these standards (or beliefs) are “realistic” and apply similarly outside the laboratory. In contrast to this overall compatibility of fairness standards, the convergence toward maximal punishment after rejections seems particularly surprising. Further research may investigate how frequency and extent of punishment depend on the efficiency of the punishment technology and on the possibility to punish agents also after acceptance.

Finally, let us emphasize the relevance of controlling for subject heterogeneity. We did so non-parametrically, by finite mixture modeling, which strikingly improved the goodness-of-fit and resolved the (counter-intuitive) insignificance of state variables

such as time and previous acceptance decision with respect to the decisions of the “representative” subject. There simply is no representative subject, and in general, analyses assuming otherwise risk getting misled. In our case, accounting for heterogeneity recovered the state variable effects, and amongst others it allowed us to pin down why demands are increasing in time—although exogenous inputs are insignificant for the representative subject. From a more general point of view, this shows that modeling heterogeneity is important not only for technical econometric reasons, but also to understand how individual behavior relates to the overall patterns of the data. The approach of finite mixture modeling, which has yet to gain wide application (but see Bardsley and Moffatt, 2007, Conte et al., 2008, and Harrison and Rutström, 2009), seems particularly helpful in this respect.

Appendix A Finite mixture modeling

A detailed review of finite mixture modeling can be found in Peel and MacLahlan (2000). The following intends to provide a brief introduction with all the information needed to follow the above analysis. The basic idea is to model a (heterogenous) population with a multiplicity of discrete player types. The main alternative approach would be to assume that subject heterogeneity can be represented using parametric distributions of random effects on the intercept (as in mixed effects models) or on other model parameters (as in say mixed logit models). The finite mixture model is more flexible in describing interdependence of parameters across player types, but it works well only if the actual distribution of types is indeed clustered in the parameter space, which then allows us to approximate it using discrete types. In our analyses, the distinction of discrete types did improve the log-likelihood strikingly, and hence we concluded that it works well.

To define the finite mixture model formally, consider a population with $K \in \mathbb{N}$ types. Each type $k \leq K$ is defined by P parameters (e.g. coefficients of a tobit or logit model), denoted as $\alpha_k := \{\alpha_{k,p}\}_{p \leq P}$. The aggregate parameter profile is $\alpha := \{\alpha_k\}_{k \leq K}$. The prior probability that a subject is of type k is denoted as ρ_k , for all $k \leq K$; the posterior follows from Bayes’ Rule as shown below. Now, let $o_{s,t}$ denote the action of subject $s \leq S$ in round $t \leq T$ of the experiment, and let $\Pr(o_{s,t}|\alpha_k)$ denote the probability/density that a player of type k (i.e. with parameters α_k) chooses $o_{s,t}$.

The log-likelihood of the model, represented by its parameters (α, ρ) , given the data set $o = (o_{s,t})$ is

$$LL(\alpha, \rho | o) = \sum_{s \leq S} \ln \sum_{k \leq K} \rho_k \cdot \prod_{t \leq T} \Pr(o_{s,t} | \alpha_k). \quad (4)$$

In turn, the posterior probability that subjects s is of type k , conditional on the actions o_s , is

$$\Pr(\text{Type} = k | o_s) = \frac{\rho_k \cdot \prod_{t \leq T} \Pr(o_{s,t} | \alpha_k)}{\sum_{k' \leq K} \rho_{k'} \cdot \prod_{t \leq T} \Pr(o_{s,t} | \alpha_{k'})}. \quad (5)$$

For the maximum likelihood estimates of the parameter profile (α, ρ) , the aggregate posterior share of type k equates with its prior share. For every individual, the posterior classification is probabilistic (although it may be rather decisive, assigning probabilities close to either 0 or 1 for all types to every subject). The posteriors for our subjects are listed in the supplementary material.

We obtained the parameter estimates by maximizing the full likelihood function jointly over all parameters (to obtain efficient estimates, see Arcidiacono and Jones, 2003). We used the Nelder-Mead algorithm as implemented in GNU R. The standard errors were derived from the information matrix, which in turn was computed numerically. Model selection was based on the Bayes information criterion (BIC, Schwarz, 1978). In our case, the BIC of a given model can be defined as

$$BIC = -LL + (K * P + K - 1) / 2 * \ln(S * T), \quad (6)$$

where $K * P + K - 1$ is the number of parameters and $S * T$ is the number of observations.

References

- Andreoni, J. and Miller, J. (1993). Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *The Economic Journal*, 103(418):570–585.
- Arcidiacono, P. and Jones, J. (2003). Finite mixture distributions, sequential likelihood and the em algorithm. *Econometrica*, 71(3):933–946.
- Bardsley, N. and Moffatt, P. (2007). The experimetrics of public goods: Inferring motivations from contributions. *Theory and Decision*, 62(2):161–193.

- Camerer, C. and Weigelt, K. (1988). Experimental tests of a sequential equilibrium reputation model. *Econometrica*, 56(1):1–36.
- Cochard, F., Nguyen Van, P., and Willinger, M. (2004). Trusting behavior in a repeated investment game. *Journal of Economic Behavior & Organization*, 55(1):31–44.
- Conte, A., Hey, J., and Moffatt, P. (2008). Mixture models of choice under risk. *Journal of Econometrics* (forthcoming).
- Cooper, D., Garvin, S., and Kagel, J. (1997). Signalling and adaptive learning in an entry limit pricing game. *The Rand Journal of Economics*, 28(4):662–683.
- Crawford, V. and Iriberri, N. (2007). Level- k auctions: Can a nonequilibrium model of strategic thinking explain the winner’s curse and overbidding in private-value auctions? *Econometrica*, 75(6):1721–1770.
- Egas, M. and Riedl, A. (2008). The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society B*, 275(1637):871.
- Fehr, E. and Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994.
- Fehr, E. and Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868):137–140.
- Fehr, E., Gächter, S., and Kirchsteiger, G. (1997). Reciprocity as a Contract Enforcement Device: Experimental Evidence. *Econometrica*, 65(4):833–860.
- Fehr, E. and Rockenbach, B. (2003). Detrimental effects of sanctions on human altruism. *Nature*, 422(6928):137–140.
- Fehr, E. and Schmidt, K. (1999). A theory of fairness, competition, and cooperation. *Quarterly journal of Economics*, 114(3):817–868.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.
- Fischbacher, U. and Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *The American Economic Review*, 100(1):541–556.

- Harrison, G. and Rutström, E. (2009). Expected utility theory and prospect theory: One wedding and a decent funeral. *Experimental Economics*, 12(2):133–158.
- Ho, T., Camerer, C., and Weigelt, K. (1998). Iterated dominance and iterated best response in experimental "*p*-beauty contests". *American Economic Review*, 88(4):947–969.
- Jung, Y., Kagel, J., and Levin, D. (1994). On the existence of predatory pricing: An experimental study of reputation and entry deterrence in the chain-store game. *The RAND Journal of Economics*, 25(1):72–93.
- Kreps, D. and Wilson, R. (1982). Reputation and imperfect information. *Journal of Economic Theory*, 27(2):253–279.
- Kübler, D. and Weizsäcker, G. (2004). Limited depth of reasoning and failure of cascade formation in the laboratory. *Review of Economic Studies*, 71(2):425–441.
- Loomes, G., Starmer, C., and Sugden, R. (2003). Do Anomalies Disappear in Repeated Markets?*. *The Economic Journal*, 113(486):C153–C166.
- Masclot, D., Noussair, C., and Villeval, M. (2010). Threat and Punishment in Public Good Experiments. *Working paper*.
- Mason, C. and Nowell, C. (1998). An experimental analysis of subgame perfect play: the entry deterrence game. *Journal of Economic Behavior and Organization*, 37(4):443–462.
- Milgrom, P. and Roberts, J. (1982). Predation, reputation and entry deterrence. *Journal of Economic Theory*, 27(2):280–312.
- Muller, L., Sefton, M., Steinberg, R., and Vesterlund, L. (2008). Strategic behavior and learning in repeated voluntary contribution experiments. *Journal of Economic Behavior & Organization*, 67(3-4):782–793.
- Neral, J. and Ochs, J. (1992). The sequential equilibrium theory of reputation building: A further test. *Econometrica*, 60(5):1151–1169.
- Neugebauer, T., Perote, J., Schmidt, U., and Loos, M. (2009). Selfish-biased conditional cooperation: On the decline of contributions in repeated public goods experiments. *Journal of Economic Psychology*, 30(1):52–60.

- Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, 92(1-2):91–112.
- Nikiforakis, N. and Normann, H. (2008). A comparative statics analysis of punishment in public-good experiments. *Experimental Economics*, 11(4):358–369.
- Peel, D. and MacLahlan, G. (2000). *Finite Mixture Models*. Wiley interscience.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464.
- Selten, R. and Stoecker, R. (1986). End behavior in sequences of finite Prisoner's Dilemma supergames A learning theory approach. *Journal of Economic Behavior & Organization*, 7(1):47–70.
- Stahl, D. (1996). Boundedly rational rule learning in a guessing game. *Games and Economic Behavior*, 16(2):303–330.

Supplementary material

Extortion in the laboratory

Overview

The supplementary material provides translations of both the instructions and the control questionnaire used in the experiment (the originals, in German, are available from the authors) as well as various figures and tables with the content described next.

- Figures 1–4 plot all demands and acceptance decisions for all P - A pairs. For each plot, the header displays the unique identifiers of P and A in the respective match (the same identifiers are used in the tables following below), and the plot contains a connected line representing the various demands of P and for each demand it displays a symbol representing either acceptance or rejection of A (a triangle pointing upwards indicates acceptance, and a little cross indicates rejection).
- Table 1 provides a tabular representation of all observations. Each “cell” in this table consists of two lines and lists the actions of a particular P - A pair. The order of P - A pairs is the same as in Figures 1–4. The upper line in each cell shows the identifier of the respective P and all demands she made. The lower line shows the identifier of A and his corresponding acceptance decisions (“1” is acceptance, “0” is rejection).
- Table 2 provides the posterior classifications for all P based on their actions.
- Table 3 provides the posterior classifications for all A based on their actions. The entries are ordered such that the first P listed in Table 2 (“S16”) is the one who was matched with the first A listed in Table 3 (“S11”), the second P (“S17”) was matched with the second A (“S12”), and so on. The posterior classifications are probabilistic, as described in Appendix A of the paper.

Experimental instructions

Instructions

You are participating in an experiment on decision making. The computer will match you randomly in groups of two. You remain in the same group for the duration of the whole experiment. One member of each group will be referred to as *A*-participant and the other member is referred to as *B*-participant. These assignments are random, too, and held constant for the duration of the experiment.

The experiment consists of 15 stages. The procedure in each stage is identical to that in all other stages. In each stage, you earn points that will be converted into Euro and paid to you after the end of the experiment. Here, the sum of the points earned in all stages is relevant. For each point, you obtain 1 cent. The payment can be collected a week after the experiment from the secretary of the chair.

The procedure of single stages The endowment of the *A*-participant is 160 points and that of the *B*-participant is 40 points. First, the *B*-participant may choose a number of points that he demands from the *A*-participant. Second, the *A*-participant has to decide whether to comply with the demand or not.

- If the *A*-participant accepts the demand, then the respective number of points is transferred to the *B*-participant.
- If the *A*-participant rejects the demand, then the *B*-participant has the option to destroy an arbitrary number of points (maximally 160) of the *A*-participant. Exercising this option bears costs of 10 points for the *B*-participant.

Then the stage ends and the resulting scores are saved for both participants. If the 15th stage is not reached yet, then a new stage begins (with constant assignments of roles).

Example 1 The *B*-participant demands x_1 points and the *A*-participants accepts the demand. The resulting scores for this stage are as follows.

A-participant: $160 - x_1$

B-participant: $40 + x_1$

Example 2 The B -participant demands x_2 points, the A -participant rejects the demand, and the B -participant chooses not to exercise the option of destroying points. The resulting scores for this stage are as follows.

$$A\text{-participant: } 160 \qquad B\text{-participant: } 40$$

Example 3 The B -participant demands x_3 points, the A -participant rejects the demand, and the B -participant destroys y_3 points of the A -participant. The resulting scores for this stage are as follows.

$$A\text{-participant: } 160 - y_3 \qquad B\text{-participant: } 40 - 10$$

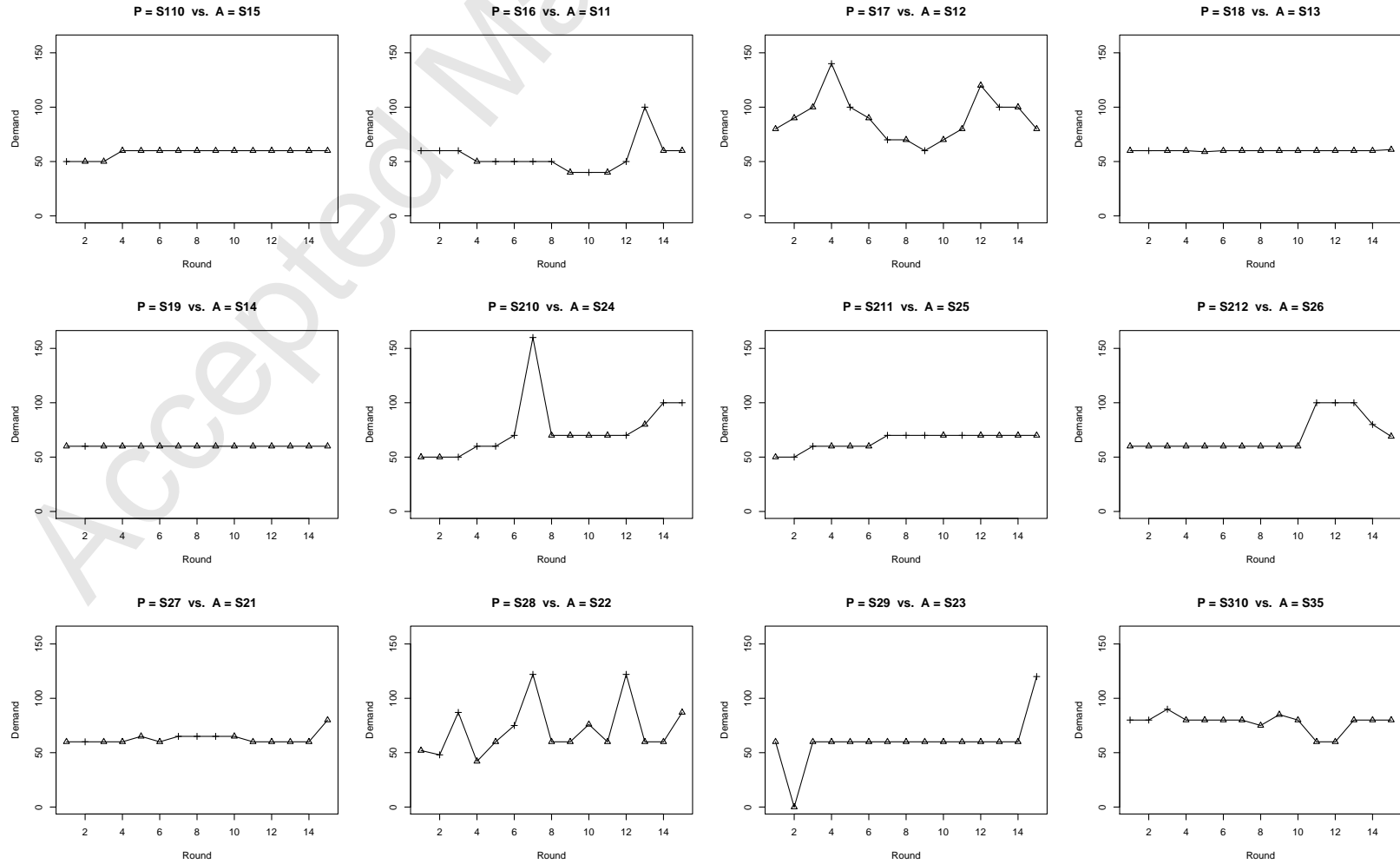
Questionnaire

Please tick the correct answers.

1. The person that is matched with me in a group
 - changes after each stage.
 - is always the same.
 - may change, but may also be the same.
2. The number of stages overall is
 - equal to 1.
 - finite, but random and not known in the beginning.
 - equal to 15.
3. The assignments of A - and B -participants in a group
 - is held constant in all stages.
 - changes after each stage.
 - may change, but may also be kept constant.
4. May payment in this experiment

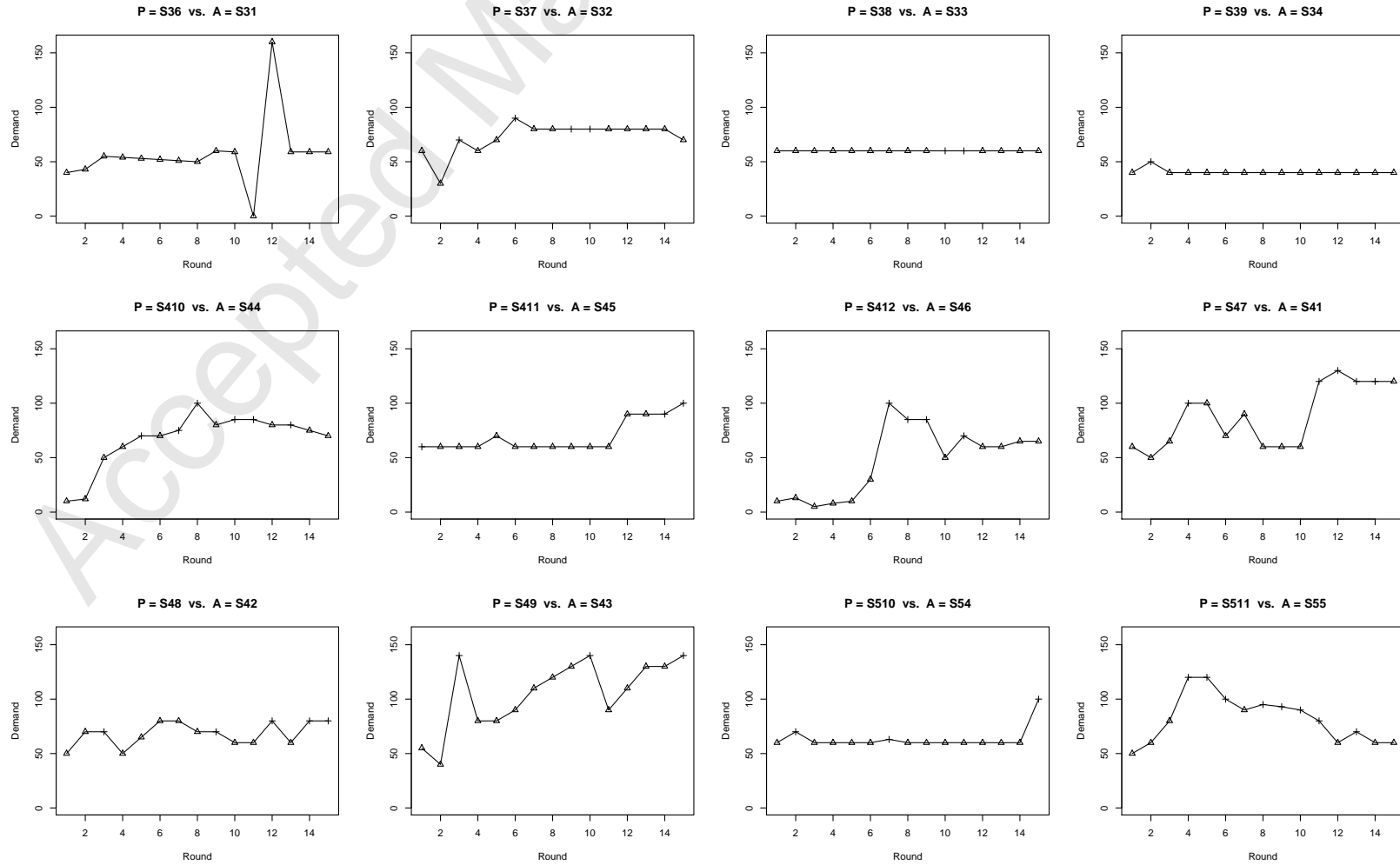
- is independent of the actions chosen.
 - results from the aggregate score of all stages.
 - results from the score earned in a randomly chosen stage.
5. If the *A*-participant rejects the demand of the *B*-participant, then
- the endowments (160 points and 40 points, respectively) remain unchanged and the current stage ends.
 - the current stage ends and both participants score 0 points in this stage.
 - the endowments (160 points and 40 points, respectively) remain unchanged and the *B*-participant is provided with an option to destroy points.
6. Exercising the option of the *B*-participant to destroy points
- bears no costs and allows him to destroy an arbitrary number of points of the *A*-participant.
 - costs 10 points and allows him to destroy an arbitrary number of points of the *A*-participant.
 - costs 10 points and allows him to destroy exactly 80 points of the *A*-participant.

Figure 1: Individual demands and acceptance (triangle)/rejection (+) decisions



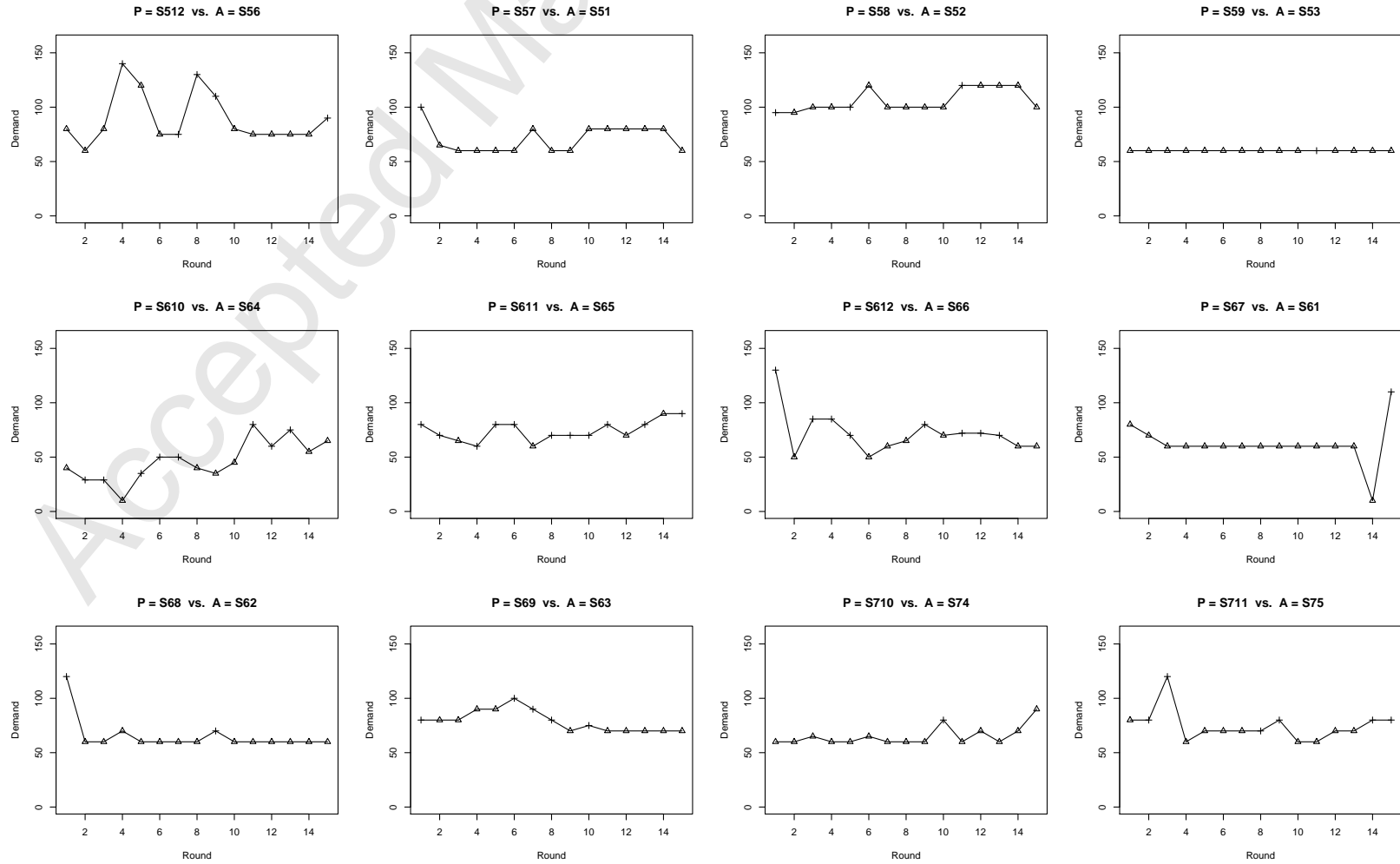
5

Figure 2: Individual demands and acceptance (triangle)/rejection (+) decisions (continued)



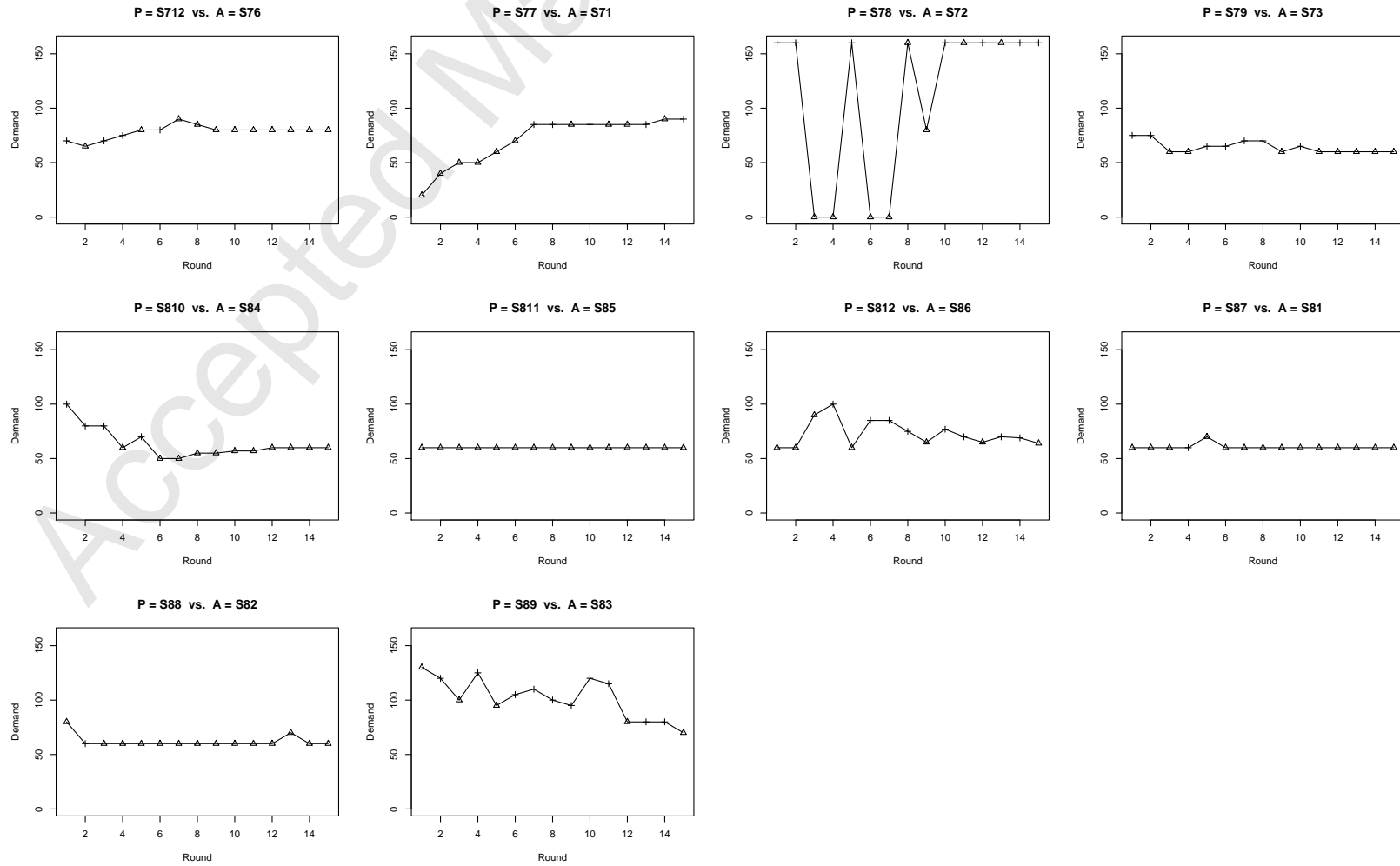
9

Figure 3: Individual demands and acceptance (triangle)/rejection (+) decisions (continued)



7

Figure 4: Individual demands and acceptance (triangle)/rejection (+) decisions (continued)



8

Table 1: All observations

Subject	Rounds														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$P = S110$	50	50	50	60	60	60	60	60	60	60	60	60	60	60	60
$A = S15$	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$P = S16$	60	60	60	50	50	50	50	50	40	40	40	50	100	60	60
$A = S11$	0	0	0	1	0	0	0	0	1	0	1	0	0	1	1
$P = S17$	80	90	100	140	100	90	70	70	60	70	80	120	100	100	80
$A = S12$	1	1	1	0	0	1	0	1	0	1	1	1	0	1	1
$P = S18$	60	60	60	60	59	60	60	60	60	60	60	60	60	60	61
$A = S13$	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
$P = S19$	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60
$A = S14$	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
$P = S210$	50	50	50	60	60	70	160	70	70	70	70	70	80	100	100
$A = S24$	1	1	0	0	0	0	0	1	1	1	1	1	0	1	0
$P = S211$	50	50	60	60	60	60	70	70	70	70	70	70	70	70	70
$A = S25$	1	0	0	1	1	1	0	0	0	1	0	1	1	1	1
$P = S212$	60	60	60	60	60	60	60	60	60	60	100	100	100	80	69
$A = S26$	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1
$P = S27$	60	60	60	60	65	60	65	65	65	65	60	60	60	60	80
$A = S21$	1	0	1	1	1	1	0	0	0	1	1	1	1	1	1
$P = S28$	52	48	87	42	60	75	122	60	60	76	60	122	60	60	87
$A = S22$	1	0	0	1	1	0	0	1	1	1	1	0	1	1	1
$P = S29$	60	0	60	60	60	60	60	60	60	60	60	60	60	60	120
$A = S23$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
$P = S310$	80	80	90	80	80	80	80	75	85	80	60	60	80	80	80
$A = S35$	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1
$P = S36$	40	43	55	54	53	52	51	50	60	59	0	160	59	59	59
$A = S31$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$P = S37$	60	30	70	60	70	90	80	80	80	80	80	80	80	80	70
$A = S32$	1	1	0	1	1	0	1	1	0	0	1	1	1	1	1
$P = S38$	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60
$A = S33$	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1
$P = S39$	40	50	40	40	40	40	40	40	40	40	40	40	40	40	40
$A = S34$	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
$P = S410$	10	12	50	60	70	70	75	100	80	85	85	80	80	75	70
$A = S44$	1	1	1	1	0	1	0	0	1	0	0	1	0	1	1
$P = S411$	60	60	60	60	70	60	60	60	60	60	60	90	90	90	100
$A = S45$	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0
$P = S412$	10	13	5	8	10	30	100	85	85	50	70	60	60	65	65
$A = S46$	1	1	1	1	1	1	0	0	0	1	0	1	1	1	1
$P = S47$	60	50	65	100	100	70	90	60	60	60	120	130	120	120	120
$A = S41$	1	1	1	0	1	1	1	1	1	1	0	0	0	0	1
$P = S48$	50	70	70	50	65	80	80	70	70	60	60	80	60	80	80
$A = S42$	1	1	0	1	1	1	1	1	0	1	1	0	1	0	0

continued on next page

Subject	Rounds														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$P = S49$	55	40	140	80	80	90	110	120	130	140	90	110	130	130	140
$A = S43$	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0
$P = S510$	60	70	60	60	60	60	63	60	60	60	60	60	60	60	100
$A = S54$	1	0	1	1	1	1	0	1	1	1	1	1	1	1	0
$P = S511$	50	60	80	120	120	100	90	95	93	90	80	60	70	60	60
$A = S55$	1	1	1	0	0	0	1	0	0	0	0	1	0	1	1
$P = S512$	80	60	80	140	120	75	75	130	110	80	75	75	75	75	90
$A = S56$	1	1	1	0	1	1	0	0	0	1	1	1	1	1	0
$P = S57$	100	65	60	60	60	60	80	60	60	80	80	80	80	80	60
$A = S51$	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$P = S58$	95	95	100	100	100	120	100	100	100	100	120	120	120	120	100
$A = S52$	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1
$P = S59$	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60
$A = S53$	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
$P = S610$	40	29	29	10	35	50	50	40	35	45	80	60	75	55	65
$A = S64$	1	0	0	1	0	0	0	1	1	1	0	0	0	1	1
$P = S611$	80	70	65	60	80	80	60	70	70	70	80	70	80	90	90
$A = S65$	0	0	1	0	0	0	1	0	0	0	0	1	0	1	0
$P = S612$	130	50	85	85	70	50	60	65	80	70	72	72	70	60	60
$A = S66$	0	1	0	0	0	1	1	1	0	1	0	0	0	1	1
$P = S67$	80	70	60	60	60	60	60	60	60	60	60	60	60	10	110
$A = S61$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
$P = S68$	120	60	60	70	60	60	60	60	70	60	60	60	60	60	60
$A = S62$	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1
$P = S69$	80	80	80	90	90	100	90	80	70	75	70	70	70	70	70
$A = S63$	0	1	1	1	1	0	0	0	1	0	1	1	1	1	1
$P = S710$	60	60	65	60	60	65	60	60	60	80	60	70	60	70	90
$A = S74$	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
$P = S711$	80	80	120	60	70	70	70	70	80	60	60	70	70	80	80
$A = S75$	1	0	0	1	1	1	1	0	0	1	1	1	1	0	0
$P = S712$	70	65	70	75	80	80	90	85	80	80	80	80	80	80	80
$A = S76$	0	1	0	0	1	0	1	1	1	1	1	1	1	1	1
$P = S77$	20	40	50	50	60	70	85	85	85	85	85	85	85	90	90
$A = S71$	1	1	1	1	1	1	0	0	1	0	1	1	0	1	0
$P = S78$	160	160	0	0	160	0	0	160	80	160	160	160	160	160	160
$A = S72$	0	0	1	1	0	1	1	1	1	0	1	0	1	0	0
$P = S79$	75	75	60	60	65	65	70	70	60	65	60	60	60	60	60
$A = S73$	0	0	1	1	0	0	0	0	1	0	1	1	1	1	1
$P = S810$	100	80	80	60	70	50	50	55	55	57	57	60	60	60	60
$A = S84$	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1
$P = S811$	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60
$A = S85$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$P = S812$	60	60	90	100	60	85	85	75	65	77	70	65	70	69	64

continued on next page

Subject	Rounds														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<i>A = S86</i>	1	1	1	0	1	0	0	0	1	0	0	1	0	0	1
<i>P = S87</i>	60	60	60	60	70	60	60	60	60	60	60	60	60	60	60
<i>A = S81</i>	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
<i>P = S88</i>	80	60	60	60	60	60	60	60	60	60	60	60	70	60	60
<i>A = S82</i>	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>P = S89</i>	130	120	100	125	95	105	110	100	95	120	115	80	80	80	70
<i>A = S83</i>	1	0	1	0	1	0	0	0	0	0	0	1	0	0	1

Table 2: Posterior classification of principals

Subject	Types			Subject	Types		
	Aggressive	Fair	Adaptive		Aggressive	Fair	Adaptive
S16	0.0219	0	0.9781	S58	4e-04	0	0.9996
S17	0.9853	0	0.0147	S59	0	1	0
S18	0	1	0	S510	0	0	1
S19	0	1	0	S511	0.0018	0	0.9982
S110	0	0	1	S512	1	0	0
S27	0	0	1	S67	1	0	0
S28	1	0	0	S68	1e-04	0	0.9999
S29	1	0	0	S69	0	0	1
S210	1	0	0	S610	6e-04	0	0.9994
S211	0	0	1	S611	0	0	1
S212	1e-04	0	0.9999	S612	0.9908	0	0.0092
S36	1	0	0	S77	0	0	1
S37	0.0014	0	0.9986	S78	1	0	0
S38	0	1	0	S79	0	0	1
S39	0	0	1	S710	0	0	1
S310	0	0	1	S711	0.7969	0	0.2031
S47	1	0	0	S712	0	0	1
S48	0	0	1	S87	0	0	1
S49	1	0	0	S88	0	0	1
S410	1e-04	0	0.9999	S89	0.0207	0	0.9793
S411	0	0	1	S810	0	0	1
S412	0.9511	0	0.0489	S811	0	1	0
S57	2e-04	0	0.9998	S812	5e-04	0	0.9995

Table 3: Posterior classification of agents

Subject	Types			Subject	Types		
	Aggressive	Fair	Adaptive		Aggressive	Fair	Adaptive
S11	0	1	0	S52	1	0	0
S12	0.9982	0	0.0018	S53	0.1282	0	0.8718
S13	0.6353	0	0.3647	S54	0.0129	1e-04	0.987
S14	0.5884	0	0.4116	S55	0	0	0.9999
S15	0.7651	0	0.2349	S56	0.0174	3e-04	0.9823
S21	0.0386	8e-04	0.9606	S61	0.0271	0	0.9729
S22	0.1729	0.001	0.8261	S62	0.1106	0	0.8894
S23	0.1383	0	0.8617	S63	0.0469	1e-04	0.953
S24	0	0.2127	0.7873	S64	0	0.9999	1e-04
S25	0.0051	0.3444	0.6505	S65	0	0.9968	0.0032
S26	0	0	1	S66	0	0.0772	0.9227
S31	1	0	0	S71	0.0014	0	0.9986
S32	0.0239	5e-04	0.9756	S72	1	0	0
S33	0.0143	0	0.9857	S73	1e-04	0.8956	0.1043
S34	0.4323	0.001	0.5667	S74	0.2504	0	0.7496
S35	0.8879	0	0.1121	S75	0	2e-04	0.9998
S41	0.0966	0	0.9034	S76	0.9703	2e-04	0.0295
S42	0	0	1	S81	0.5998	0	0.4002
S43	1	0	0	S82	0.8233	0	0.1766
S44	1e-04	1e-04	0.9997	S83	0	0.0363	0.9637
S45	0.0528	0	0.9472	S84	0.2317	0.0236	0.7447
S46	0.0013	0.0038	0.9949	S85	0.3773	0	0.6227
S51	0.8698	0	0.1302	S86	0	2e-04	0.9998