

Güth, Werner; Otsubo, Hironori

**Working Paper**

## Whom to blame? An experiment of collective harming and punishing

Jena Economic Research Papers, No. 2011,046

**Provided in Cooperation with:**

Max Planck Institute of Economics

*Suggested Citation:* Güth, Werner; Otsubo, Hironori (2011) : Whom to blame? An experiment of collective harming and punishing, Jena Economic Research Papers, No. 2011,046, Friedrich Schiller University Jena and Max Planck Institute of Economics, Jena

This Version is available at:

<https://hdl.handle.net/10419/56916>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



# JENA ECONOMIC RESEARCH PAPERS



# 2011 – 046

## **Whom to blame? An experiment of collective harming and punishing**

by

**Werner Güth  
Hironori Otsubo**

[www.jenecon.de](http://www.jenecon.de)

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact [markus.pasche@uni-jena.de](mailto:markus.pasche@uni-jena.de).

Impressum:

Friedrich Schiller University Jena  
Carl-Zeiss-Str. 3  
D-07743 Jena  
[www.uni-jena.de](http://www.uni-jena.de)

Max Planck Institute of Economics  
Kahlaische Str. 10  
D-07745 Jena  
[www.econ.mpg.de](http://www.econ.mpg.de)

© by the author.

# Whom to blame? An experiment of collective harming and punishing\*

Werner Güth<sup>†</sup>

*Max Planck Institute of Economics*

Hironori Otsubo<sup>‡</sup>

*Soka University*

October 12, 2011

## Abstract

In a situation, where it is efficient for one of two parties to pollute but highly inefficient if both parties do so, the harmed third party can freely impose its damage claims on both parties what crucially determines which equilibrium to expect. Whereas “equality before the law” requires equal punishments, efficient equilibria are predicted when holding one party responsible and letting the other escape punishment. After discussing equilibrium selection for the game, we report on an experiment with two treatment variables: one that determines when the harmed party announces its potential damage claims and one varying a game- and justice-unrelated difference between the two culprits. According to our experimental data, “equality before the law” dominates but is weakened by asymmetry in wealth and the possibility to announce sanctions.

Keywords: Law and economics; Equality vs. efficiency; Equilibrium selection; Laboratory experiments

JEL Classification: C72, C92, K00

---

\*Generous financial support from the Max Planck Society is gratefully acknowledged. The authors would like to thank Rico Löbel for his excellent assistance in data collection.

<sup>†</sup>Address: Strategic Interaction Group, Max Planck Institute of Economics, Kahlaische Straße 10, 07745 Jena, Germany; Tel.: +49-3641-686-620; Fax: +49-3641-686-623; E-mail: [gueth@econ.mpg.de](mailto:gueth@econ.mpg.de).

<sup>‡</sup>Corresponding author. Address: Faculty of Economics, Soka University, 1-236 Tangi-machi, Hachioji-shi, Tokyo, 192-8577, Japan; Tel.: +81-42-691-9316; E-mail: [hiroecon@gmail.com](mailto:hiroecon@gmail.com).

# 1 Introduction

In the law and economics literature (e.g., Cooter and Ulen, 2011; Posner, 2011), it is claimed that law has been designed and possibly (as legal standards) evolved to promote efficiency. Such efficiency enhancing legal systems may, however, violate other legal principles as, for instance, “equality before the law.” The paper introduces a highly stylized situation where these two principles, efficiency enhancement and equality before the law, are conflicting. More specifically, we consider the following situation:

- There are two agents who may pollute an environment. When only one pollutes, the polluter gains and nobody else suffers. When both agents pollute, the environment is damaged, and neither of them gains whereas a third agent suffers from the damage.
- This third agent can punish either agent more or less so as to reclaim the damage *before* or *independently of* the two agents’ decision to pollute.

Clearly, the symmetry of the two agents and the symmetry of their collective responsibility in causing the damage implies that “equality before the law” means to punish them equally. However, this practice is detrimental for efficiency: there exists two efficient strict equilibria which are, however, asymmetric so that a symmetry invariant, but inefficient mixed strategy equilibrium is likely to result. By declaring in advance to hold only one of them responsible for the damage the third party may induce the agent, who would be fully responsible for the damage, to abstain from polluting and the other agent to pollute. It is thus efficiency enhancing to violate “equality before the law.”

What factors may guide the third agent’s decision in such a situation? We have two factors in mind. The first concerns the choice timing of the third agent. If, as we have already discussed, the third agent can arbitrarily discriminate between the two agents by preannouncing her irrevocable commitment on how to fine the two agents, the third agent may be able to induce a favorable outcome. This precommitment

strategy may be in vain when the two agents cannot hear the third agent's preannouncement. The other factor is an inequality in wealth between the two agents. This spurious, game- and justice-unrelated difference between the two agents may suggest a punishment schedule that holds the wealthier agent more responsible for the damage and could avoid the mixed strategy equilibrium.

We formulate a stylized game model of this situation, derive equilibrium predictions by applying the theory of equilibrium selection (Harsanyi and Selten, 1988), and conduct a laboratory experiment designed to address the following research questions:

- Is the third agent more likely to choose a biased punishment schedule when the two agents are able to observe the third agent's choice than when they are unable?
- Is the third agent more likely to choose a biased punishment schedule when the two agents differ in wealth than when they are equally wealthy?

To address these questions we designed and implemented a two-by-two factorial experimental design that varies the choice timing of the third agent (sequential vs. simultaneous) and the endowments of the two agents (symmetric vs. asymmetric).

In section 2, we introduce the game models formally and then investigate and select among their equilibria by applying the theory of equilibrium selection. Section 3 describes the experimental design. After analyzing the experimental data in section 4, we conclude in section 5.

## 2 Theory

### 2.1 Setup

The game has three players,  $X$ ,  $Y$ , and  $Z$ , and consists of two stages in sequence. In the first stage, player  $X$  chooses a number  $p \in S_X = [0, d]$ . In the subsequent stage, the remaining players,  $Y$  and  $Z$ , simultaneously choose  $s_i \in S_i = \{R, B\}$ ,  $i \in \{Y, Z\}$ .

	<i>Z</i>	<i>R</i>	<i>B</i>
<i>Y</i>		<i>R</i>	<i>B</i>
		<i>R</i>	<i>B</i>
		<i>R</i>	<i>B</i>
		<i>B</i>	<i>B</i>
		<i>B</i>	<i>B</i>

Table 1: The trimatrix subgame after  $X$ 's choice of  $p \in [0, d]$  with players  $Y$  and  $Z$  and payoffs listed in alphabetic order

Table 1 describes the payoff matrix between active players  $Y$  and  $Z$ , given player  $X$ 's choice. Each cell lists the three players' payoffs in alphabetical order. If both  $X$  and  $Y$  choose  $R$ , each of the three players earns a payoff of  $e$ . If  $X$  and  $Y$  choose differently, the one whose choice is  $B$  earns  $e + b$  while the other two players earn  $e$  each. If both  $X$  and  $Y$  choose  $B$ , players  $X$ ,  $Y$ , and  $Z$  earn  $e - c$ ,  $e - p$ , and  $e - (d - p)$ , respectively. Here,  $d$  is the damage claim of  $X$ , in case of  $Y$  and  $Z$  choosing  $B$ , for which  $X$  has full discretion in holding  $Y$  and  $Z$  responsible via his choice of  $p$ . We assume that  $b > 0$ ,  $c > 0$ ,  $d > 0$ , and  $e > 0$ . Note that once  $(B, B)$  has been chosen by  $Y$  and  $Z$ , the choice of  $p$  is payoff irrelevant for  $X$ . In our view, this allows for two interpretations of role  $X$ , the one reclaiming the damage cost  $d$  and the one of judge being confronted with the task to assign damage payments.

## 2.2 The Sequential Order Protocol

Under the sequential order protocol, players  $Y$  and  $Z$  simultaneously choose either  $R$  or  $B$  in the second stage, knowing  $X$ 's choice in the first stage (See Figure 1 (a)). Thus, the analysis begins with the second stage.

– Insert Figure 1 about here –

For any  $p \in (0, d)$ , neither  $(s_Y, s_Z) = (R, R)$  nor  $(s_Y, s_Z) = (B, B)$  is an equilibrium whereas both  $(s_Y, s_Z) = (R, B)$  and  $(s_Y, s_Z) = (B, R)$  are strict equilibria. There also exists an additional properly mixed equilibrium for  $0 < p < d$ . Denote by  $\sigma_i = (\sigma_i(R), \sigma_i(B))$ ,  $i \in \{Y, Z\}$ , player  $i$ 's mixed strategy in which  $i$  uses  $R$  with probability  $\sigma_i(R)$  and  $B$  with probability  $\sigma_i(B)$ . Then, the weak mixed-strategy equilibrium is  $(\sigma_Y, \sigma_Z) = \left( \left( \frac{d-p}{b+d-p}, \frac{b}{b+d-p} \right), \left( \frac{p}{b+p}, \frac{b}{b+p} \right) \right)$ . In case of  $p = 0$ ,

the only strict equilibrium is  $(s_Y, s_Z) = (B, R)$  whereas the only weak equilibrium is  $(s_Y, s_Z) = (R, B)$ . In case of  $p = d$ , the only strict equilibrium is  $(s_Y, s_Z) = (R, B)$  whereas the only weak equilibrium is  $(s_Y, s_Z) = (B, R)$ .

For  $p \in \{0, d\}$ , one thus can avoid the ambiguity of the equilibrium concept by postulating strictness what, however, would not work for  $p = \frac{d}{2}$ . Here one might require symmetry invariance (i.e., the solution should not depend on naming players and strategies). This rules out both strict equilibria and leaves only the symmetry invariant mixed-strategy equilibrium.

**Proposition 1.** *Imposing strictness in case of  $p \in \{0, d\}$  and equilibrium selection in case of  $p \in (0, d)$  yields a unique benchmark solution, namely:*

- $(s_Y, s_Z) = (B, R)$  for  $0 \leq p < \frac{d}{2}$ ,
- $(s_Y, s_Z) = (R, B)$  for  $\frac{d}{2} < p \leq d$ , and
- $(\sigma_Y, \sigma_Z) = \left( \left( \frac{d-p}{b+d-p}, \frac{b}{b+d-p} \right), \left( \frac{p}{b+p}, \frac{b}{b+p} \right) \right)$  with  $p = \frac{d}{2}$ .

For  $p \in (0, d)$  and  $p \neq \frac{d}{2}$ , the equilibria  $(B, R)$  and  $(R, B)$  remain to be strict but are not symmetric so that equilibrium selection theory (Harsanyi and Selten, 1988) can be applied. Due to best reply and isomorphic invariance as well as monotonicity the risk dominant solution is  $(R, B)$  if  $p > \frac{d}{2}$  and  $(B, R)$  if  $p < \frac{d}{2}$ .<sup>1</sup> For the mixed strategy equilibria  $(\sigma_Y, \sigma_Z)$ , any  $p \in (0, d)$  implies positive probability for player  $X$ 's loss event  $(B, B)$ . Player  $X$  should therefore avoid choosing  $p \in (0, d)$  but is indifferent between  $p = 0$  and  $p = d$ . Applying again symmetry invariance makes  $X$  choose  $p = 0$  and  $p = d$  with probability  $\frac{1}{2}$  each as a correlated equilibrium.

### 2.3 The Simultaneous Order Protocol

In a sharp contrast to the sequential order protocol, the simultaneous order protocol rules out that any player has been informed about another player's choice when deciding, i.e., all three players decide independently (see Figure 1 (b)). This may be due to:

---

<sup>1</sup>Isomorphic invariance implies symmetric invariance.

- player  $X$  deciding before  $Y$  and  $Z$  who are, however, not informed about  $X$ 's choice of  $p$  when choosing, or
- all three players deciding simultaneously.

Although all three players decide independently, player  $X$  seems as free in his choice of  $p \in [0, d]$  as in the sequential order protocol. This is due to the fact that the optimal value of  $p$  for player  $X$  does not depend at all on how players  $Y$  and  $Z$  decide. More specifically, for all (pure or mixed) strategy combinations of  $Y$  and  $Z$ , player  $X$ 's payoff does not depend at all on his choice of  $p$ . In the terminology of Harsanyi and Selten (1988) the game with the player set  $\{X\}$ , containing only player  $X$  as an active player, is a *cell*. This is a subset of the player set that is closed with respect to the best reply correspondence, i.e., the best reply sets of cell players depend only on the behavior of the other cell players but not at all on that of non-cell players.

**Proposition 2.** *According to the simultaneous order protocol,  $\{X\}$  is a cell.*

What Proposition 2 implies is that the equilibrium choices of  $Y$  and  $Z$  do not affect what is optimal for  $X$ . In other words: in any equilibrium of the simultaneous order protocol, players  $Y$  and  $Z$  react to  $X$ 's choice of  $p$  as analyzed in the sequential order protocol.

However, since all  $p \in [0, d]$  are best replies by  $X$ , it still remains ambiguous which  $p$  will be chosen by  $X$ . In the terminology of Harsanyi and Selten (1988) the cell solution is ambiguous due to the pathology of the cell game.

**Proposition 3.** *According to the simultaneous order protocol, any value of  $p$  defines an equilibrium set in which this  $p$  is combined with the combinations of  $Y$ 's and  $Z$ 's choices as derived for the sequential order protocol, i.e., for  $p \in \{0, d\}$  there are two pure strategy equilibria and for  $p \in (0, d)$  there exist three equilibria, two of which are in pure strategies of player  $Y$  and  $Z$  and one in mixed strategies.*

To overcome this troubling ambiguity, let us assume that player  $X$  cares for resolving strategic uncertainty by applying the same equilibrium selection criteria



as applied for the sequential order protocol. Thus, player  $X$  relies on strictness when this yields uniqueness and, when not, applies equilibrium selection. This obviously rules out any  $p \in (0, d)$  as optimal for player  $X$  if

- $p \in (0, d)$  results in the mixed-strategy equilibrium

$$(\sigma_Y, \sigma_Z) = \left( \left( \frac{d-p}{b+d-p}, \frac{b}{b+d-p} \right), \left( \frac{p}{b+p}, \frac{b}{b+p} \right) \right)$$

for  $Y$  and  $Z$  and thus assigns positive probability to the loss event  $(s_Y, s_Z) = (B, B)$  for player  $X$ ,

- rules out at least  $p = \frac{d}{2}$  when relying on equilibrium selection since by symmetry invariance  $(\sigma_Y, \sigma_Z)$  would be the solution for  $p = \frac{d}{2}$ ,
- whereas in case of  $p \in \{0, d\}$  and the strictness requirement the probability of this loss event for  $X$  is zero.

**Proposition 4.** *Assuming that player  $X$  relies on the same equilibrium selection as applied for the sequential order protocol,  $X$  will choose either  $p = 0$  to induce  $(s_Y, s_Z) = (B, R)$  or  $p = d$  to induce  $(s_Y, s_Z) = (R, B)$  and use either choice with probability  $\frac{1}{2}$ .*

This has largely reduced the set of possible benchmark solutions. Like “Buridan’s donkey,” facing two equally good food items, we should not worry too much how  $X$  will decide in case of indifference and impose symmetry invariance meaning that  $X$  applies  $p = 0$  and  $p = d$  with equal probability. Although in different ways, we have derived the same outcome predictions for the sequential and simultaneous order protocols. According to our benchmarks, one thus would expect invariance across the protocol treatments.

### 3 Experiment

The four treatments differ from one another in terms of the order protocol (sequential vs. simultaneous) and the show-up fees (symmetric vs. asymmetric). They are

referred to as *Seq-Sym* (sequential order protocol and symmetric show-up fees), *Seq-Asym* (sequential order protocol and asymmetric show-up fees), *Sim-Sym* (simultaneous order protocol and symmetric show-up fees), and *Sim-Asym* (simultaneous order protocol and asymmetric show-up fees), respectively.

A total of two hundred seven student subjects from various fields of study at the Friedrich-Schiller University of Jena were recruited via the ORSEE software (Greiner, 2004). We organized two sessions per treatment and invited twenty seven subjects per session.<sup>2</sup> No subject was allowed to participate in more than one session. All eight sessions were conducted in the experimental laboratory of the Max Planck Institute of Economics in Jena, Germany, with thirty-two PCs connected in a network. The experiment was programmed and conducted with the software *z-Tree* (Fischbacher, 2007). A session lasted about 60 minutes, including reading instructions and paying subjects.

Upon arrival at the laboratory, subjects were randomly assigned to computer terminals separated from one another by partitions. Any form of communication between subjects was strictly forbidden throughout the session, and questions were answered individually by the experimenter. After all subjects being seated, they were asked to read written instructions silently at their own pace. Once all of them indicated readiness for the experiment, the experimenter read the instructions aloud so that all information became common knowledge. Then, the subjects were given six control questions designed to check their understanding of the instructions.<sup>3</sup>

Each session consists of only three rounds (iterations). Prior to the first round, the computer randomly formed three groups of nine subjects each. Group composition remained the same so that no interaction between groups took place throughout the session. Then, for each group the computer randomly assigned three subjects to the role of *X*, another three subjects to the role of *Y*, and the remaining three subjects to the role of *Z*. They were labeled “*X*,” “*Y*,” and “*Z*,” respectively. Subjects

---

<sup>2</sup>Since we were unable to recruit twenty seven subjects for the second session of treatment *Sim-Sym*, we decided to run this session with eighteen subjects instead.

<sup>3</sup>The English instructions and control questions for treatment *Seq-Asym* are available in Appendix A.

$Y \backslash Z$	$R$	$B$
$R$	15, 15, 15	15, 15, 30
$B$	15, 30, 15	10, 15 - $p$ , 5 + $p$

Table 2: Payoffs in the experiment

retained their role throughout the session.

The sequence of each round was identically structured in all treatments. At the beginning of a round, each subject was randomly matched with two other subjects in her group who were assigned the opposite roles. We implemented a perfect stranger design ensuring that subjects never play against other subjects more than once. Each round consists of two stages. In the first stage, only subjects in the role of  $X$  submitted their decisions. They were asked to choose one of the eleven different combinations of ten balls. These combinations differed from one another in terms of the numbers of black and white balls.<sup>4</sup> For example, choosing a combination that consists of more black balls than white balls means punishing  $Y$  more than  $Z$ . In the second stage, the remaining subjects (i.e., subjects in the roles of  $Y$  and  $Z$ ) simultaneously choose either  $R$  or  $B$  with knowledge of  $X$ 's decisions in the sequential order protocol and without that knowledge in the simultaneous order protocol. At the end of a round, each subject was informed of her decision as well as the decisions of the other two subjects and how much she earned in the round.

At the end of a session, only one round was selected for payment as follows; the experimenter randomly chose a volunteer subject, who was asked to draw one ball from a box containing three balls labeled 1 through 3. The drawn number determined the round for payment.

In the experiment, we set  $b = 15$ ,  $c = 5$ ,  $d = 10$ , and  $e = 15$ , respectively, and therefore payoffs were determined according to the trimatrix shown in Table 2. Here  $p$  denotes the number of black balls in the combination chosen by  $X$ . Values in each entry are denoted in "points."

---

<sup>4</sup>See Appendix A.

Treatment	Round	Number of black balls										
		0	1	2	3	4	5	6	7	8	9	10
<i>Seq-Sym</i>	1	2	1	0	0	1	9	0	2	2	0	1
	2	3	1	1	0	1	8	1	0	0	1	2
	3	3	1	1	0	0	8	1	0	0	1	3
<i>Seq-Asym</i>	1	2	2	1	5	4	2	0	0	0	1	1
	2	4	1	1	1	2	3	1	2	1	1	1
	3	5	1	1	3	1	2	1	0	1	2	1
<i>Sim-Sym</i>	1	0	0	0	0	1	10	3	0	0	0	1
	2	0	0	1	1	2	8	1	0	0	1	1
	3	0	0	0	0	0	11	1	1	1	0	1
<i>Sim-Asym</i>	1	3	1	1	3	0	6	2	1	1	0	0
	2	3	1	2	3	2	4	0	0	0	1	2
	3	5	1	0	1	2	6	0	0	1	1	1

Table 3: Frequencies of the number of black balls in the combinations chosen by  $X$  subjects

## 4 Results

### 4.1 Behavior in the First Round

#### 4.1.1 $X$ Subjects' Behavior

We begin the Results section by comparing the observed choices of subjects in the role of  $X$  in the first round with the unique benchmark solution. Hereafter, subjects in the roles of  $X$ ,  $Y$ , and  $Z$  are called  $X$  subjects,  $Y$  subjects, and  $Z$  subjects, respectively.

Table 3 presents the observed distribution of the combinations chosen by  $X$  subjects by treatment and round. Particularly interesting is their behavior in the first round as no elements of learning and experience come into play. There are two major discernible features in the results. First,  $X$  subjects failed to fully blame one subject (i.e., zero or ten blackballs) in all treatments. Instead, they predominantly declared intermediate punishment schedules. These results are clearly inconsistent with Propositions 1 and 4. Second, subjects showed a tendency to seek equal punishments (i.e., five blackballs) when the show-up fees are symmetric while they were more likely to blame the wealthier  $Z$  subjects when the show-up fees are asymmetric.

Comparison		Result	$p$ -value
<i>Seq-Sym</i>	vs. <i>Seq-Asym</i>	<i>s</i>	0.02
<i>Sim-Sym</i>	vs. <i>Sim-Asym</i>	<i>ns</i>	0.0776
<i>Seq-Sym</i>	vs. <i>Sim-Sym</i>	<i>ns</i>	0.7065
<i>Seq-Asym</i>	vs. <i>Sim-Asym</i>	<i>ns</i>	0.4322

Table 4: Results of the Wilcoxon rank sum test. *s* and *ns* stand for significance and non-significance at the 0.05 level, respectively

These tendencies remained unchanged, regardless of which order protocol was implemented. “Equality before the law” seems to be a strong motivation of punishing behavior by  $X$  subjects, especially in case of symmetric subjects.

**Result 1:** “Equality before the law” in the sense of  $p = \frac{d}{2}$  (i.e., 5 black balls in Table 3) dominates, especially in the case of symmetry.

Did the observed distributions of punishment schedules differ significantly across the four treatments? Figure 2 shows four pairwise comparisons of empirical distributions of punishment schedules in the four treatments. The figure indicates that the distributions of the two treatments *Seq-Sym* and *Seq-Asym* look least similar. To validate this statistically, we performed the Wilcoxon rank sum test.<sup>5</sup> The results are summarized in Table 3: only the comparison between *Seq-Sym* and *Seq-Asym* reveals a significant difference ( $p$ -value  $< 0.05$ ). Although subjects seem to have behaved differently between *Sim-Sym* and *Sim-Asym*, we barely failed to reject the null hypothesis of no difference at the 0.05 level. In the remaining two comparisons, the same null hypothesis was soundly accepted ( $p$ -value  $> 0.15$ ).

– Insert Figure 2 about here –

#### 4.1.2 Behavior Facing Punishment

We have just seen that  $X$  subjects failed to choose the punishment schedules predicted by theory; they declared more equal punishments when  $Y$  and  $Z$  subjects

<sup>5</sup>In the first round, there were fifteen independent observations in *Sim-Sym* and eighteen independent observations in the remaining three treatments.

were symmetric in wealth and sought more unequal punishments favoring poorer subjects (i.e.,  $Y$  subjects) when they were not. Then, how did  $Y$  and  $Z$  participants behave in the second stage?

– **Insert Figure 3 about here** –

We first take a look at  $Y$  and  $Z$  subjects' behavior when the order protocol was simultaneous. In the two treatments with the simultaneous order protocol,  $Y$  and  $Z$  subjects had to make their decision without knowing  $X$ 's behavior in the first stage. Figure 3 displays observed relative frequency distributions of the second stage outcomes in the first round. A comparison between the *Sim-Sym* and *Sim-Asym* treatments shows that the proportion of the outcome  $(B, R)$  is higher in *Sim-Asym* (38.3%) than in *Sim-Sym* (33.3%) whereas the proportion of the outcome  $(R, B)$  is higher in *Sim-Sym* (33.3%) than in *Sim-Asym* (27.8%). Did  $Z$  subjects reveal an intrinsic compensation incentive when they were wealthier than  $Y$  subjects? The ratio of the  $(B, R)$  frequency to the  $(R, B)$  frequency for *Sim-Sym* is 1 whereas the corresponding value for *Sim-Asym* is 1.4. The Fisher's exact tests to examine whether there is any significant difference between the two respective distributions of the second stage outcomes, however, failed to reject the null hypothesis of no difference ( $p$ -value=0.2257).

How did  $Y$  and  $Z$  subjects respond to  $X$ 's behavior when the order protocol was sequential? Recall that the frequency of  $X$  subjects' choice is distributed (almost) symmetrically around its mode at equal punishments in *Seq-Sym* and skewed toward more severe punishment on  $Z$  subjects in *Seq-Asym* (see Table 3). Such discrepancies seem to help  $Y$  and  $Z$  subjects coordinate on the outcome  $(B, R)$ , i.e., the efficient outcome that favors  $Y$  subjects. The relative frequencies of the two efficient outcomes  $(B, R)$  and  $(R, B)$  are 44.4% and 27.8% in *Seq-Sym* whereas the corresponding values are 55.6% and 5.6% in *Seq-Asym*. The fact that  $(B, R)$  was chosen more frequently than  $(R, B)$  in *Seq-Sym* can hardly be explained by its first stage outcomes that favored neither  $Y$  nor  $Z$  subjects. In *Seq-Asym*, however,  $X$  subjects' behavior that sanctioned  $Z$  subjects more may be responsible for the

significant difference in the relative frequencies of the two efficient outcomes in the second stage.

**Result 2:** *There is weak evidence that asymmetry in wealth led  $Y$  and  $Z$  subjects to the second stage outcome  $(B, R)$  that favors the poorer  $Y$  subjects when the first stage behavior was not observable. When their choice was observable, asymmetry in wealth induced  $X$  subjects to sanction the richer  $Z$  subjects more in the first stage, which in turn made the outcome  $(B, R)$  more appealing in the second stage.*

## 4.2 $X$ Subjects' Behavior over Three Rounds

Despite the small number of repetitions, repeated play of the same game may boost subjects' understanding of the game. Did repetition of the same game help  $X$  subjects to realize that it was efficient to blame only one of the two agents?

To answer this question, we computed the average number of black balls, chosen by  $X$  subjects, separately for each treatment. The results are plotted in Figure 4 with the dashed line indicating equal punishment. The figure clearly indicates that the answer to the question is negative: there is no discernible time trend toward punishing only one of the  $Y$  and  $Z$  subjects. Rather  $X$  subjects, on average, remained indecisive in designating who would be fully responsible for the potential damage. An interesting observation is that the tendency of  $X$  subjects to blame richer subjects remained constant across the three rounds of successive play.

– Insert Figure 4 about here –

**Result 3:** *On average,  $X$  subjects persistently relied on “equality before the law” across rounds when  $Y$  and  $Z$  subjects were equally wealthy and on punishing the poorer  $Y$  subjects less when they differed in wealth.*

What Result 3 suggests is the so-called “deep pocket phenomenon”: judges as well as agents react to a game- and justice-unrelated payoff asymmetry by trying to limit payoff asymmetry across agents. This may be very debatable from a doctrinal point

<i>Seq-Sym</i>	Total	Success	Failure	<i>Seq-Asym</i>	Total	Success	Failure
<i>E</i>	9	6	3	<i>E</i>	2	1	1
<i>U<sub>Y</sub></i>	4	1	3	<i>U<sub>Y</sub></i>	14	6	8
<i>U<sub>Z</sub></i>	5	0	5	<i>U<sub>Z</sub></i>	2	1	1

<i>Sim-Sym</i>	Total	Success	Failure	<i>Sim-Asym</i>	Total	Success	Failure
<i>E</i>	10	8	2	<i>E</i>	6	3	3
<i>U<sub>Y</sub></i>	1	0	1	<i>U<sub>Y</sub></i>	8	7	1
<i>U<sub>Z</sub></i>	4	1	3	<i>U<sub>Z</sub></i>	4	1	3

Table 5: Persistency of  $X$  subjects' behavior over three rounds by treatment and type

of view but appears to be a matter of fact in many countries.<sup>6</sup>

Finally, one can explore the individual dynamics of  $X$  subjects. Did those who sought equal punishments in round 1 continue to punish equally later? Similarly, did those who chose unequal punishments in round 1 continue to “rule” in the same way later? To more closely look at the individual dynamics of  $X$  subjects, they are classified into the following three types based on their first round behavior: Type  $E$  is the group of those who sought equal punishments (i.e., five black balls), Type  $U_Y$  is the group of those who chose unequal punishments in favor of  $Y$  subjects (i.e., less than five black balls), and Type  $U_Z$  is the group of those who chose unequal punishments in favor of  $Z$  subjects (i.e., more than five black balls). Then, for each of the three types we count the number of  $X$  subjects who punished the same way over all three rounds.

Table 5 summarizes the results. In the table, “Success” counts if an  $X$  subject of a type continued to punish in the same way over three rounds and “Failure” if not. Thus, the number of successes indicates persistency of  $X$  subjects' behavior over three rounds. An observation being worth noting is that Type  $E$  subjects in the *Sym* treatments were more successful in punishing in the same way over three rounds than those in the *Asym* treatments. The number of  $X$  subjects who persistently chose equal punishments is six out of the nine subjects (67%) in treatment *Seq-Sym*,

<sup>6</sup>Since most labor court disputes are settled by pre-trial agreements, statistical evidence for the “deep pocket phenomenon” is, to the best of our knowledge, poor and mostly anecdotal.



one out of the two (50%) in *Seq-Asym*, eight out of the ten (80%) in *Sim-Sym*, and three out of the six (50%) in *Sim-Asym*.

Another interesting observation is that those who punished unequally (i.e., Types  $U_Y$  and  $U_Z$  subjects) struggled to stay in the same punishment style. The number of  $X$  subjects who persistently chose unequal punishments is one out of nine subjects (11%) in treatment *Seq-Sym*, seven out of sixteen (44%) in *Seq-Asym*, one out of five (20%) in *Sim-Sym*, and eight out of twelve (67%) in *Sim-Asym*.

**Result 4:** *X* subjects punishing equally in round 1 more likely maintained their punishment style in the *Sym* treatments than in the *Asym* treatments. Moreover, *X* subjects punishing unequally in round 1 were more successful in maintaining the same punishment style in the *Asym* treatments than in the *Sym* treatments.

Altogether our results justify the conclusion that “equality before the law” in the sense of  $X$  subjects choosing  $p = \frac{d}{2}$ , i.e., equal punishments, is an obvious and appealing principle which can be weakened by game- and justice-unrelated asymmetries that, except for *Seq-Asym*, does not question its modality: choosing the combination with five blackballs is the most frequent  $X$  subjects’ choice across rounds and treatments except for *Seq-Asym*, where it is never the most frequent choice.

## 5 Concluding Remarks

Legal doctrines may be conflicting if not hierarchically ordered. Here we concentrated on “equality before the law” in the sense that equally responsible parties should be equally sanctioned and on “efficiency enhancement” in the sense of joint payoff or welfare maximization and designed a game for which these two legal principles suggest different “rulings” or punishment schedules. After deriving the game theoretic benchmark solution which does not depend on whether agents know or do not know the punishment schedule, we have experimentally implemented the game in a two-by-two factorial design (sequential vs. independent, symmetric vs. asymmetric endowments).

Our main findings are:

- a clear dominance of “equality before the law,” and
- this dominance is only significantly weaker in case of asymmetric endowments and sequential decision making.

These findings suggest that “judges” ( $X$  subjects) but also the two other agents themselves are trying to balance agents’ payoffs by punishing the richer agent more (deep pocket phenomenon). That the parties ( $Y$  and  $Z$  subjects) themselves try to reduce payoff inequalities, e.g. due to inequ(al)ity aversion (Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999) had to be expected. It is, however, much more debatable when judges or courts engage in active redistribution. If the rich cannot succeed in court, this could imply that the weaker party is much less inclined to respect its contractual obligations, e.g. employees by shirking, and that the stronger party will avoid appealing to courts. As a matter of fact the “deep pocket phenomenon,” which seems so typical for labor court settlements, may finally induce a legal subculture where local practices evolve without much connection to the legal system and thus without democratic control.

## References

- Bolton, G. E., Ockenfels, A. 2000. ERC: A theory of equity, reciprocity, and competition. *American Economic Review* 90, 166–193.
- Cooter, R., Ulen, T. 2011. *Law and Economics*, 6<sup>th</sup> edition. New Jersey: Prentice Hall.
- Fehr, E., Schmidt, K. M. 1999. A theory of fairness, competition and cooperation. *Quarterly Journal of Economics* 114, 817–868.
- Fischbacher, U. 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10, 171–178.
- Greiner, B. 2004. The online recruitment system ORSEE 2.0—A guide for the organization of experiments in economics. Working paper, University of Cologne.
- Güth, W., Kliemt, H., Ockenfels, A. 2001. Retributive Responses. *Journal of Conflict Resolution* 45, 453–469.
- Harsanyi, J. C., Selten, R. 1988. *A General Theory of Equilibrium Selection in Games*. Cambridge: MIT Press.
- Posner, A. R. 2011. *Economic Analysis of Law*, 8<sup>th</sup> edition. New York: Aspen Publishers.
- Siegel, S., Castellan, N. J. 1988. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.

## Appendix

### A Instructions and Control Questions for Treatment *Seq-Asym* (originally written in German)

#### INTERACTIVE DECISION MAKING EXPERIMENT SUBJECT INSTRUCTIONS

##### Introduction

Welcome! You are about to participate in an interactive decision making experiment funded by the Max Planck Institute of Economics. If you have a mobile phone, please switch it off now.

Please read the instructions carefully. Your decisions, as well as the decisions of the other participants, will determine your payoff according to the rules that will be explained shortly.

Please note that hereafter any form of communication between the participants is strictly prohibited. If you have any questions, please raise your hand. The experimenter will come to assist you.

##### Detailed Information on the Experiment

This experiment is fully computerized. You will be making your decisions by clicking on appropriate buttons on the screen. All the participants are reading the same instructions and taking part in this experiment for the first time, as you are.

A total of 27 persons are participating in this experiment. At the beginning of the experiment, the computer will randomly assign nine participants to the role of **X**, other nine participants to the role of **Y**, and the remaining nine participants to the role of **Z**. Therefore, there are nine Xs, nine Ys, and nine Zs. Your role will remain the same throughout the experiment and so will the roles of the other participants.

The experiment consists of **three rounds**. In each round, you will be matched with two other participants who are assigned to different roles. For example, if you are X, then you will be matched with one Y and one Z every round. **The composition of your group will change every round.** This means that your group members will be different from one round to the next. You have no chance of interacting with the same participants more than once.

In the experiment, we will use “points” as the currency. **At the end of the experiment, one of the three rounds will be selected for payment by the procedure that will be explained at the end of the instructions.** The points you have earned in the selected round will be converted to euro at the rate of **1 point = €0.50** and paid to you in cash.

In addition, each participant will be paid a show-up fee for having shown up on time. **The show-up fee differs across the three roles.** You will receive a **€5.00** show-up fee if you are X, a **€2.50** show-up fee if you are Y, and a **€7.50** show-up fee if you are Z.

### Description of the Task

Each round consists of **two stages**:

**Stage 1:** X will be asked to choose one of the 11 different combinations of **black** and **white** balls shown in the table below. In each combination, the upper number represents the number of black ball(s) and the lower number represents the number of white ball(s). As you see, each combination contains exactly **10** balls.

	<b>Combinations</b>										
Black	0	1	2	3	4	5	6	7	8	9	10
White	10	9	8	7	6	5	4	3	2	1	0

**Stage 2:** At the beginning of Stage 2, Y and Z will be informed which combination has been chosen by X in Stage 1. In other words, both Y and Z will know the numbers of black and white balls in that combination. Then, Y and Z will be asked

to *simultaneously* choose either **Red** or **Blue**. In other words, Y and Z will choose one of the two alternatives without knowing the other's decision.

### How to Compute Payoffs

After completion of Stage 2, the computer will automatically compute your payoff as well as the other group members' payoffs for the current round according to the following rules:

1. If both Y and Z chose Red in Stage 2,

X's payoff = 15 points
Y's payoff = 15 points
Z's payoff = 15 points

regardless of the combination chosen by X in Stage 1.

2. If Y chose Red and Z chooses Blue in Stage 2,

X's payoff = 15 points
Y's payoff = 15 points
Z's payoff = 30 points

regardless of the combination chosen by X in Stage 1.

3. If Y chose Blue and Z chooses Red in Stage 2,

X's payoff = 15 points
Y's payoff = 30 points
Z's payoff = 15 points

regardless of the combination chosen by X in Stage 1.

4. If both Y and Z chose Blue in Stage 2,

X's payoff = 10 points

Y's payoff = 15 – number of black ball(s) in the combination points

Z's payoff = 15 – number of white ball(s) in the combination points

Below are examples illustrating how to compute payoffs:

**Example 1:** In Stage 1, X chose a combination with 9 black balls and 1 white ball. In Stage 2, both Y and Z chose Red. Then, each of them will earn 15 points.

**Example 2:** In Stage 1, X chose a combination with 4 black balls and 6 white balls. In Stage 2, Y chose Red whereas Z chose Blue. Then, X and Y will earn 15 points each whereas Z will earn 30 points.

**Example 3:** In Stage 1, X chose a combination with 7 black balls and 3 white balls. In Stage 2, Y chose Blue whereas Z chose Red. Then, X and Z will earn 15 points each whereas Y will earn 30 points.

**Example 4:** In Stage 1, X chose a combination with 2 black balls and 8 white balls. In Stage 2, both Y and Z chose Blue. Then, X will earn 10 points, Y will earn 13 (= 15 – 2) points, and Z will earn 7 (= 15 – 8) points.

### **Feedback Information at the End of Each Round**

After completion of Stage 2, the computer will exhibit a results screen that shows

- your decision as well as the decisions of the other two members in your group, and
- your payoff for the current round.

### **End of the Experiment**

After completing the experiment, the computer will display a history screen that presents your payoffs (in points) in the three rounds. Then, the experimenter will

determine a seat number by drawing one chip from a box that contains 27 chips labeled 1 through 27. The participant with the selected seat number will be a *volunteer*. The volunteer participant will be asked to come forward and then draw one ball from a box that contains three balls labeled 1 through 3. This will determine the round for payment.

A summary screen will display

- the round chosen by the volunteer,
- the points you have earned in the chosen round,
- the corresponding earnings in euros,
- your show-up fee, and
- your total earnings in euros.

Please remain at your cubicle until asked to come forward and receive payment for the experiment.

---

When you are ready for the experiment, please click on the  button on the screen. When all participants have pressed this button, the experimenter will start reading the instructions aloud. After that, you will have to answer a series of six questions designed to check your understanding of the instructions. After all participants have completed answering the six questions, the experiment will begin.

Please remember that no communication is allowed during the experiment. If you encounter any difficulties, please raise your hand. The experimenter will come to assist you.



## Control Questions

Q.1 (True/False Question) Your role will change every round.

True / False

*Answer:* False. Your role will stay the same throughout the experiment.

Q.2 (True/False Question) You may be matched with the same participants more than once.

True / False

*Answer:* False. You will never be matched with the same participants again.

Q.3 (True/False Question) Each participant will be paid a €2.50 show-up fee.

True / False

*Answer:* False. The show-up fee differs across the three roles. You will receive a **€5.00** show-up fee if you are X, a **€2.50** show-up fee if you are Y, and a **€7.50** show-up fee if you are Z.

Q.4 (Multiple Choice Question) In Stage 1, X chose a combination with 7 black balls and 3 white balls. In Stage 2, both Y and Z chose Red. How many points will they earn for this round?

- A. X will earn 10 points, Y will earn 8 points, and Z will earn 12 points.
- B. X will earn 10 points, Y will earn 12 points, and Z will earn 8 points.
- C. X will earn 15 points, Y will earn 30 points, and Z will earn 15 points.
- D. None of the above.

*Answer:* D. Since both Y and Z chose Red in Stage 2, each of the group members will earn 15 points.

Q.5 (Multiple Choice Question) In Stage 1, X chose a combination with 4 black balls and 6 white balls. In Stage 2, both Y and Z chose Blue. How many points will they earn for this round?

- A. X will earn 10 points, Y will earn 9 points, and Z will earn 11 points.
- B. X will earn 10 points, Y will earn 11 points, and Z will earn 9 points.
- C. Each of them will earn 15 points.
- D. None of the above.

*Answer:* B. In Stage 1, X chose a combination with 4 black balls and 6 white balls. Since both Y and Z chose Blue in Stage 2, X will earn 10 points, Y will earn 11 ( $= 15 - 4$ ) points, Z will earn 9 ( $= 15 - 6$ ) points.

Q.6 (Multiple Choice Question) In Stage 1, X chose a combination with 1 black ball and 9 white balls. In Stage 2, Y chose Red and Z chose Blue. How many points will they earn for this round?

- A. X will earn 10 points, Y will earn 6 points, and Z will earn 14 points.
- B. X will earn 10 points, Y will earn 14 points, and Z will earn 6 points.
- C. X will earn 15 points, Y will earn 15 points, and Z will earn 30 points.
- D. None of the above.

*Answer:* C. Since Y chose Red and Z chose Blue in Stage 2, both X and Y will earn 15 points each whereas Z will earn 30 points.

# Figures

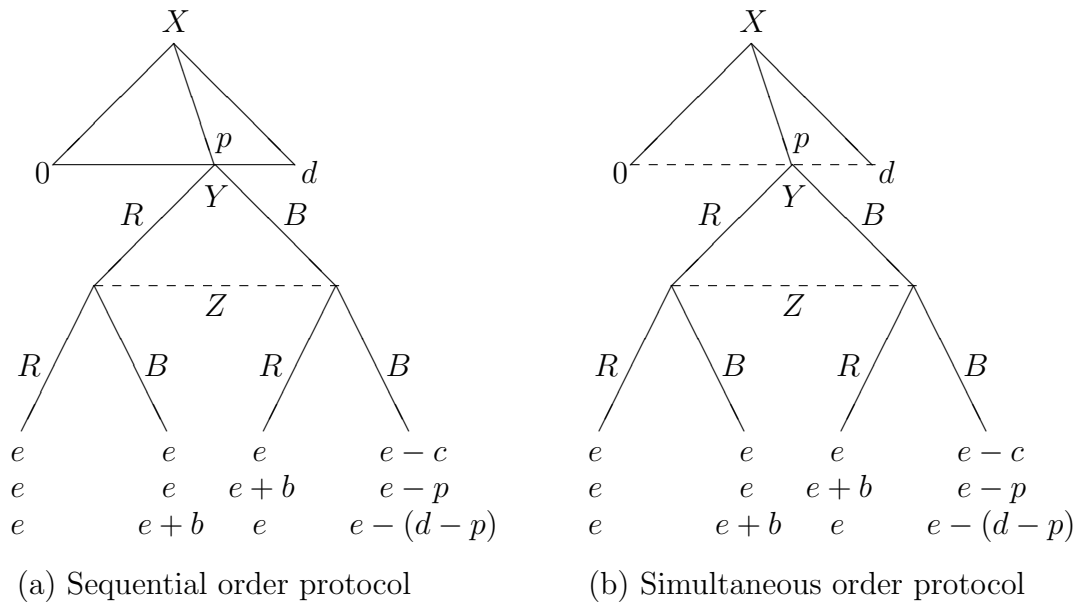


Figure 1: Game trees

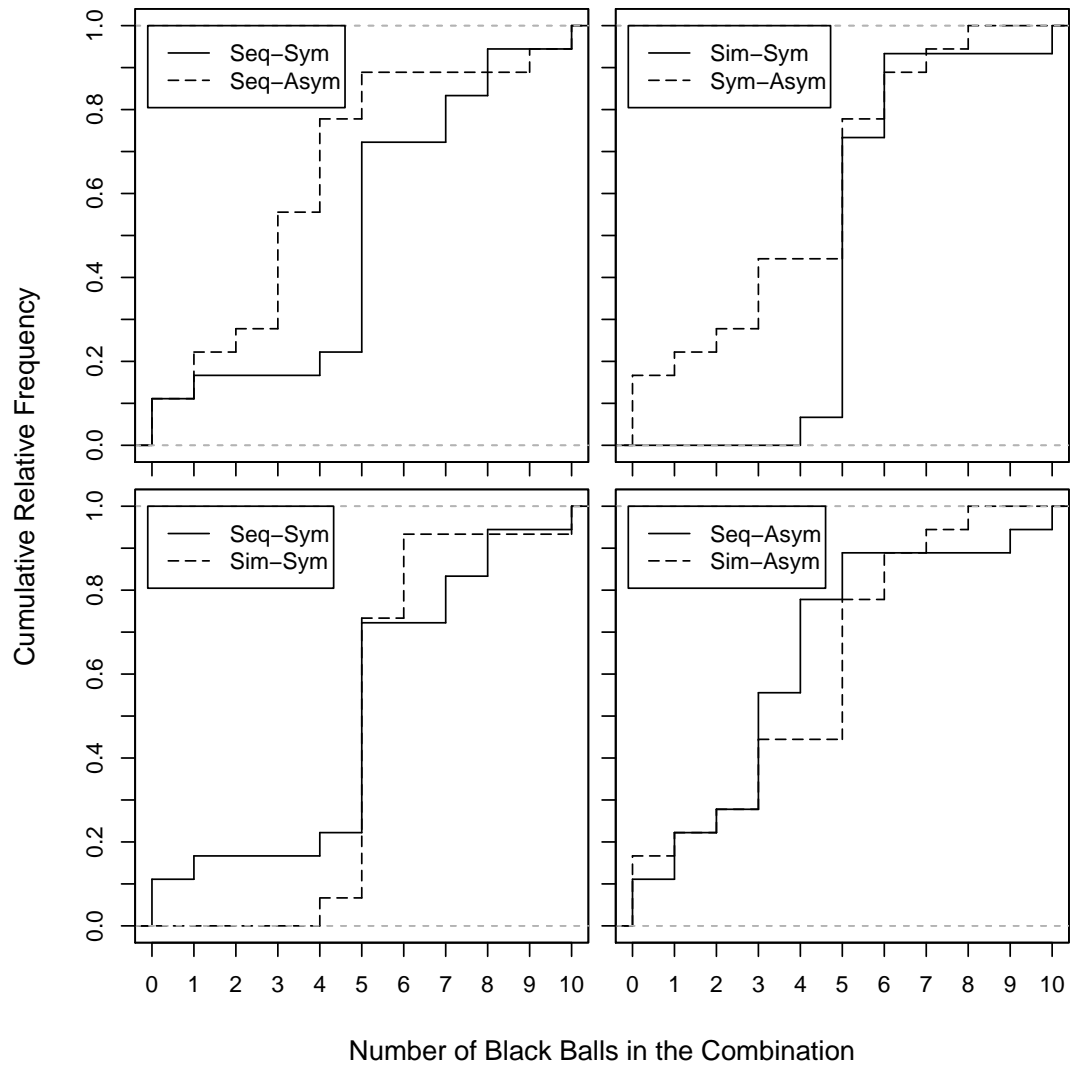


Figure 2: Pairwise comparisons of empirical distributions of the number of black balls in the combination chosen by X

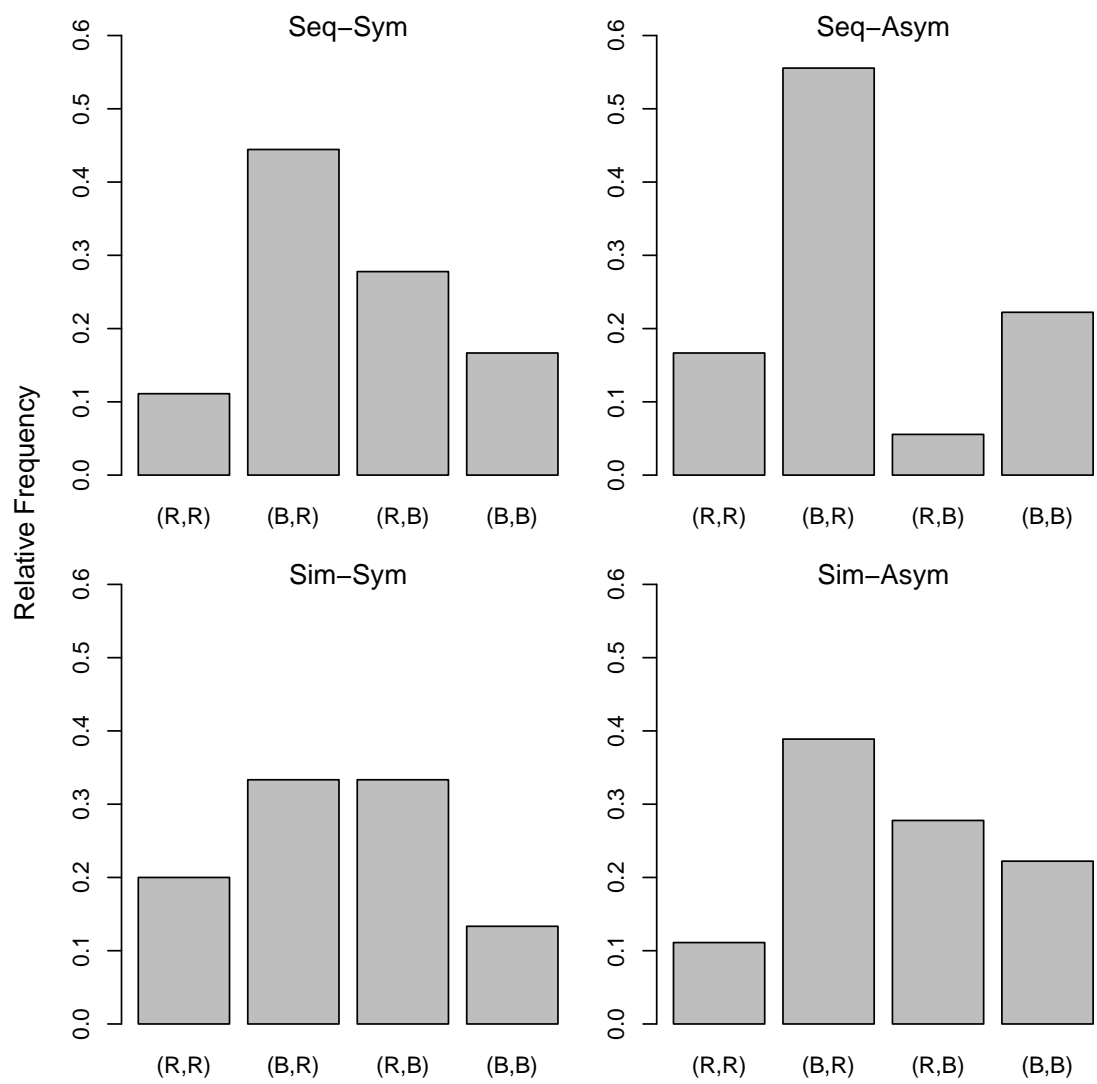


Figure 3: Observed relative frequency distributions of the second stage outcomes in the first round

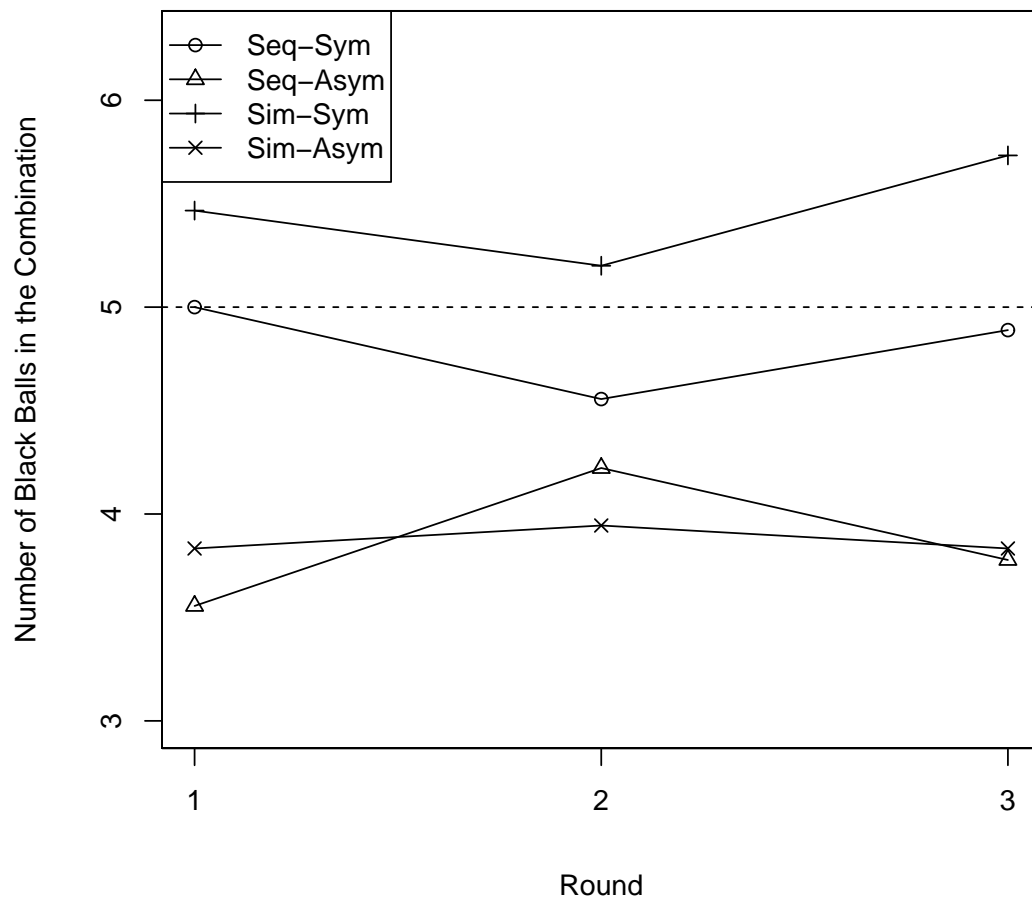


Figure 4: Time trends of the average number of black balls in the combination chosen by X