

Shift the Blame: An Experimental Analysis of a Dictator Game with Delegation and Punishment

Björn Bartling Urs Fischbacher
University of Munich University of Zurich

February 15, 2007
– first draft –

Abstract:

We elicit experimentally the attribution of responsibility for outcomes of delegated decisions. A player can take a decision between a fair and an unfair distribution of a given endowment or he can delegate this decision to another player with aligned interests. A third type of player is adversely affected when the unfair distribution is chosen and can assign costly punishment points. We find a clear pattern of punishment: mainly the player that takes the unfair choice is punished while the player that delegated the decision is spared. To distinguish between intrinsic and extrinsic motives for delegation we also conducted treatments without punishment. We find less but significant delegation without punishment. This suggests a disutility of taking the unfair decision as an intrinsic motivation for delegation, besides the extrinsic motivation of shifting the blame.

Keywords: responsibility, experiment, delegation
JEL: C91, D63, L22

1. Introduction

The delegation of decisions is an integral part of everyday work life. The management literature (e.g. Henderson 2006) singles out three main reasons for delegation. It (i) helps avoiding work overload, (ii) allows benefiting from employees' information advantages, and (iii) motivates employees and develops their skills. The focus of our paper is much different from the reasons for delegation that are discussed in this existing literature.¹

In our experimental design we abstract from issues of work overload, asymmetric information, and employee motivation and skill development. We conduct an experiment to

¹ For a formal analysis of optimal delegation within organizations see, for example, Aghion and Tirole (1997).

elicit the attribution of responsibility for outcomes of delegated decisions, and to distinguish between intrinsic and extrinsic motives for delegating a decision. When an unpleasant decision is to be taken, the disutility of taking the decision constitutes an intrinsic motivation for delegation. When those subjects that might be adversely affected by the decision have the option to punish, then the expected punishment constitutes an extrinsic motivation for delegation – at least if it is mainly the person that took the decision and not the person that delegated the decision that is punished.

Consider the following example. Taiwan's BenQ recently shut down their money-losing mobile phone manufacturing unit in Germany, formerly owned by Siemens. Critics claimed that the Siemens management spun off the mobile phone business only with intent not to file for insolvency 'in person' but to delegate that task to BenQ. Following intense public protest in Germany that stressed Siemens' responsibility for their former mobile phone unit, the Siemens management felt obliged to grant facilities to the BenQ employees that lost their jobs. In this example the presumed delegation of the liquidation of the mobile phone unit did not delegate the responsibility for this act, i.e., in the public opinion it did not shift the blame for the liquidation. In contrast to BenQ, Siemens was within reach for local protesters (not only geographically but also because of their broad product range sold in Germany), which could explain why Siemens was the main target of the protest. In our experiment the delegating subject and the delegate are symmetric and we show that responsibility can then well be shifted.

We propose a new variant of the dictator game, which involves treatment variations along two dimensions: delegation and punishment. In treatments with delegation, the dictator (player A) has the option to delegate the sharing decision to another player whose payoffs are aligned (player B). In treatments with punishment, the receivers (players C) have the option to assign costly punishment points.

In our experiment there are groups of four players. Each group consists of one player A, one player B, and two players C. In all treatment, there are two distributions of payoffs among the players. The fair distribution grants all players the same payoff; the unfair distribution grants players A and B a much higher payoff than the two players C. In the base treatment without delegation and without punishment (*noD&noP*), player A decides between the fair and the unfair distribution. Players B and C cannot take a decision in this treatment. In treatments with delegation (*D&noP* and *D&P*) player A can decide between the fair and the unfair distribution or he can delegate this decision to player B. If player A delegates then player B must decide between the fair and the unfair distribution. In treatments with

punishment (*noD&P* and *D&P*) player A or B decides between the fair and the unfair distribution. One player C is randomly chosen and can, after player A or B has taken his decision, spend one point to reduce the other players' points by up to seven points. The seven punishment points can be assigned to a single player or they can be split and assigned to two or all three other players.

We find that in treatment *noD&noP* about a third of players A choose the fair distribution. In treatment *D&noP* 17% of players A delegate the decision between the fair and the unfair distribution to player B. Across subjects, the delegating players A come out of the fraction of players A that choose the fair distribution in treatment *noD&noP*. What could be the motive for delegation? In treatment *noD&noP* some players A, despite their fair choices, might prefer the unfair distribution that yields 9 instead of 5 points – if only they did not have to take that decision themselves. Such players A might incur a disutility of being responsible for the unfair distribution and therefore choose the fair distribution. But if it is in someone else's responsibility to take the decision between the fair and the unfair distribution, then they would prefer the unfair distribution to be chosen.

The observation that 17% of players A delegate in treatment *D&noP* questions theories of social preferences (Andreoni 1989, Fehr and Schmidt 1999, Bolton and Ockenfels 2000) that do not predict delegation, and that interpret fair decisions as reflecting a preference for fair outcomes. Our result suggests that not all "fair" players A have a preference for fair outcomes but that they rather incur a disutility of not being fair. This constitutes an intrinsic motivation to delegate a decision.

What does the punishment pattern of players C look like? For players C we used the strategy method. We asked each player C (and not only the randomly chosen player C whose choice was implemented) for each possible case whether they want to punish and if yes how they want to allocate the 7 punishment points to the three other players in the group. In treatment *noD&P* there are only two possible cases: either player A is fair or unfair. In treatment *D&P* there are four possible cases: either player A does not delegate and is fair or unfair or he delegates and player B is fair or unfair.

We find that almost exclusively the player that took the unfair decision himself is punished. Only if player A delegates and B chooses the unfair distribution, A is also punished to some extent. This shows that – in contrast to the BenQ-Siemens example – responsibility can be delegated: player A is punished less severely if player B took the unfair choice as compared to the case in which player A himself chooses the unfair distribution.

Interestingly, if player A delegates and player B chooses the fair distribution, then player A is not punished. This observation is at odds with most models of intention based reciprocity (Rabin 1993, Levine 1998, Charness and Rabin 2002, Dufwenberg and Kirchsteiger 2004).² If player A delegates and player B chooses the unfair distribution then players C interpret both the move by player A and the move by player B as unfriendly: Player B has chosen the unfair distribution, and player A could have chosen the fair distribution but did not to do so. However, players C's valuation of player A's decision to delegate should not be affected by the subsequent move of player B. If delegation by player A is interpreted as unfriendly when player B subsequently chooses the unfair distribution, then delegation should also be interpreted as unfriendly as well when player B subsequently chooses the fair distribution. Theories of intention based reciprocity would thus predict that the extent of the punishment for player A in case of delegation should not depend on player B's decision. However, we find that player A is punished significantly harsher when player B is unfair than when he is fair.

Given the punishment behavior of players C, how is the choice of players A and B affected in treatments with punishment? In treatment *noD&P* two thirds of players A choose the fair distribution, which is reasonable given the punishment behavior by players C. In treatment *D&P* 55% of players A delegate the decision – compared to 17% of players A in treatment *D&noP*. Now there is an additional motive for delegation: the punishment pattern by players C that mainly punish the player that chooses the unfair distribution himself. This constitutes an extrinsic motivation to delegate a decision.

The remainder of the paper is organized as follows. Section 2 explains our experimental design in detail. Section 3 exposes our experimental predictions. Section 4 presents the results. Section 5 concludes.

2. The Experimental Design

To distinguish between different motives for delegating a decision and to elicit the attribution of responsibility for the resulting outcome, we implemented the following three-stage game. There are groups of four players. Each group consists of one player A, one player B, and two players C. Player A or player B can decide between a fair and an unfair distribution of 20 points among the four players in the group. The fair distribution assigns 5 points to each of

² It seems as if in this experiment, a person is only punished when the outcome is unfair *and* the person intentionally caused this unfair outcome – as modelled in Falk and Fischbacher (2006).

the four players; the unfair distribution assigns 9 points each to players A and B and 1 point each to players C. Each point was converted into 3 CHF (about \$ 2.4) at the end of the experiment. Table 1 shows the feasible distributions.

	Player A	Player B	Player C	Player C
Fair distribution	5	5	5	5
Unfair distribution	9	9	1	1

Table 1: Feasible distributions among which player A or player B can choose.

We considered two treatment variations – with or without *delegation* and with or without *punishment* – and conducted both treatment variations in combination.

In treatments *without delegation* player A must, in the first stage of the game, decide between the fair and the unfair distribution. In the second stage of the game, player B cannot take any decision. In treatments *with delegation* player A can, instead of taking the decision himself, delegate the decision to player B. If player A delegates in the first stage of the game, then player B must decide between the fair and the unfair distribution in the second stage of the game. Player B can neither refuse to take the decision nor delegate it to another player. If player A does not delegate, then player B cannot take any decision.

In treatments *without punishment* no player C can take a decision. In treatments *with punishment* one player C is randomly chosen. In the third stage of the game – after either player A or player B has taken the decision between the two distributions – the chosen player C can assign costly punishment points to player A, B, and the other player C. The chosen player C can spend one of his points to reduce the other players' points by up to seven points. The seven punishment points can be assigned to a single player or they can be split and assigned to two or all three other players. However, no player can be taken more points than he received in the first or second stage of the game. The chosen player C can also decide to assign less than seven reduction points and leave the unassigned points void. The player C that is not chosen cannot take any decision. Figure 1 summarizes the sequence of possible moves in the game.

2.1.4 No Delegation and No Punishment (*noD&noP*)

In treatment *noD&noP* player A cannot delegate the decision between the two distributions to player B but must, in the first stage of the game, take that decision himself. In the second stage of the game player B thus cannot take any decision. There is no punishment stage such that no player C can take a decision.

2.2 Procedural Details

Irrespective of the treatment, each game was played one-shot. All treatments were framed in a neutral manner.³ The experimental procedure was computerized with the software “z-Tree” (Fischbacher 2007). Subjects were students from the University of Zurich and the Swiss Federal Institute of Technology in Zurich. The recruitment of subjects was conducted with the software “ORSEE” (Greiner 2004). A total of 548 subjects participated in the experiment. They were randomly allocated a role as players A, B or C upon arrival at the lab, a quarter of them as players A, a quarter as players B, and a half as players C. No subject participated more than once. 144, 128, 140, and 136 subjects participated in the treatments *D&P*, *noD&P*, *D&noP*, and *noD&noP*, respectively. Sessions without punishment lasted about 45 minutes; sessions with punishment lasted about 60 minutes. On average, subjects earned CHF 25 (about \$20) in the sessions without punishment and CHF 23.25 (about \$18.50) in the sessions with punishment, which include a show-up fee of CHF 10 (about \$8).

3. Behavioral Predictions

3.1 Predictions of choices by players A and B in treatments without punishment

Consider first the basic treatment *noD&noP*, which is a variant of the dictator game. In the classic dictator game, the experimenter grants discretionary power over an amount of money to one of two anonymously matched subjects (the dictator). The dictator must decide how much of this amount to send to the other subject (the receiver) and how much to keep. The standard result is that a proportion of dictators give some amount to the receiver, even when

³ Participants were called players A, B, and C. Feasible distributions of points were called distribution 1 and distribution 2; we avoided value laden terms like “fair” and “unfair”. Players C could “reduce” other players’ points; we did not use the term “punish”.

no one (including the experimenter) can observe their choices.⁴ Sharing behavior in dictator games is largely interpreted as inequity aversion (Fehr and Schmidt 1999, Bolton and Ockenfels 2000) or altruism (Andreoni 1989).⁵ We expect similar behavior in our variant of the game, which leads to our first prediction.

Prediction 1: In treatment *noD&noP* a proportion of players A choose the fair distribution.

Consider now the treatment *D&noP*. This is a variant of a dictator game in which player A can (but need not) decide between the fair and the unfair distribution. Player A can also delegate the decision to player B who is then the dictator. Models of inequity aversion or altruism would not predict delegation in treatment *D&noP*. Player A either prefers the fair or the unfair distribution. If he delegates he has no control over player B's choice, which could thus be his less preferred distribution.⁶

However, there might be players A who do not choose the fair distribution because they want to achieve a fair outcome but because they want avoid being unfair or want avoid appearing as unfair (see also the discussion on Lazear et al. (2006) below). Such players A might prefer the unfair distribution – if only they did not have to take that decision themselves. Taking the unfair decision in person would incur a disutility of being responsible for the unfair distribution and this might lead to the choice of the fair distribution when delegation is not an option. But if it is in someone else's responsibility to take the decision between the fair and the unfair distribution, then they would prefer the unfair distribution to be chosen. Across subjects, we thus predict that out of the fraction of “fair” players A in treatment *noD&noP* some players delegate the decision in treatment *D&noP*. Similarly, we predict that across subjects a fraction of players A that choose the unfair distribution in treatment *noD&noP* delegates the decision in treatment *D&noP*. However, we hypothesize that this fraction is smaller because such players A have revealed (across subjects) that their disutility of being directly responsible for the unfair distribution is not sufficiently pronounced to keep them from choosing it. By delegating they have less to gain and, in

⁴ For an overview of the literature see Camerer (2003) or Fehr and Schmidt (2006).

⁵ In the Fehr-Schmidt model player *i*'s utility function is given by $u_i = x_i - \alpha_i / 3 \sum \max\{x_i - x_j, 0\} - \beta_i / 3 \sum \max\{x_j - x_i, 0\}$, where x denotes the players' monetary payoffs, α_i player *i*'s aversion to unfavorable inequity and β_i his aversion to favorable inequity. The division by $n-1=3$ reflects a normalization for the number of players, which is $n=4$ in our game. Player *i* as a player A thus chooses the fair distribution if $\beta_i \geq .75$.

⁶ There could be the knife-edge case that a player A is indifferent between the fair and the unfair distribution. Delegation is then also a possible move. We cannot exclude this possibility, but we consider it very unlikely that (all) players A that delegate are exactly indifferent between the two distributions and that indifference explains why we observe delegation. A simple control treatment where player A incurs a cost of ϵ in case of delegation could eliminate this explanation.

addition, incur the risk that player B chooses the fair distribution. For players A that have chosen (across subjects) the fair distribution, it does not constitute a risk that player B might choose the fair distribution as this is what they would have done without the option to delegate. This leads to our second prediction.

Prediction 2: In treatment *D&noP* a proportion of players A delegate the decision between the fair and the unfair distribution to player B. Across subjects, a higher fraction of players A that choose the fair distribution in treatment *noD&noP* delegates in treatment *D&noP* as compared to players A that choose the unfair distribution in treatment *noD&noP*.

The first part of Prediction 2 is related to findings by Dana, Cain, and Dawes (2005) and Lazear, Malmendier, and Weber (2006). They propose a variant of the dictator game in which the dictator has the option to “exit” the game. If the dictator exits, then the potential receiver receives nothing and, importantly, never gets to know that a dictator game could have been played. The dictator then receives a fraction of the amount that he otherwise could have shared with the receiver. In Dana et al. (2005) dictators had the choice to play a \$10 dictator game or to exit the game and receive \$9. They report that about 30% of dictators prefer to exit and take \$9 instead of playing the game and choosing (\$9, \$1) or (\$10, \$0). Models of other-regarding preferences cannot explain why some dictators exit the game because (\$9, \$0) is dominated by (\$9, \$1). Dana et al. (2005) conclude that giving often does not reflect a genuine concern for others’ payoffs but rather a desire not to violate others’ expectations.⁷ Lazear et al. (2006) report similar results from a dictator game with an exit option. They conclude that there are not only subjects who *like sharing* (inequity averse or altruistic subjects) and those who *dislike sharing* (purely self-interested subjects), but also subjects who *dislike not sharing*. Subjects of the latter type might feel compelled to share out of shame or guilt if they are in a situation where they are asked to share, but if they had the choice they would prefer to avoid such situations altogether.

In our game, the two players C know that they take part in a dictator game in the role of receivers, and this cannot be avoided by delegating the decision. Not wanting to disappoint the receivers’ expectations, as proposed by Dana et al. (2005), thus cannot be a motive for delegation. However, the dislike of not sharing might be avoided by delegation since this deprives oneself of the possibility to share – even though, in contrast to dictator games with

⁷ In a related paper, Dana, Weber, and Kuang (2005) show that fairness decreases substantially when the connection between choices and outcomes is obfuscated.

exit option, the sharing situation as such cannot be avoided.⁸ The dislike of not sharing, as proposed by Lazear et al. (2006), could then constitute a motive for delegation.

Consider now the behavior of players B in treatment *D&noP* when player A delegates. Models of inequity aversion or altruism would predict that the proportion of players B that chose the fair distribution is identical to the proportion of players A that chose the fair distribution in treatment *noD&noP*. There is no strategic difference in the two decisions, and the distribution of other-regarding preferences among players A and B should be identical.

However, we predict that a higher fraction of players B chooses the unfair distribution compared to the fraction of players A in treatment *noD&noP*. The reason for this prediction is twofold. First, if player A delegates the decision then player B observes that player A did not choose the fair distribution even though he could have done so. Player B thus infers from this observation that he would not be the only subject that does not choose the fair distribution and this might induce player B to opt for the unfair distribution. Second, the structure of our game introduces an implicit hierarchy. Player A is at the top of this hierarchy. He can decide between the fair and the unfair distribution or he can delegate the decision to player B. Player B can only decide when player A delegates. In this sense, player B is below player A in the hierarchy. Players C cannot decide anything in treatments without punishment. They are thus at the bottom of the hierarchy. Now, if player A delegates then player B might ponder why player A did so. If player A wanted to, he could have chosen the fair distribution. Furthermore, choosing the unfair distribution might involve a disutility, which player A wants to avoid. Player B might thus infer that player A expects him to choose the unfair distribution, and as a “willing executioner” further down in the hierarchy he might indeed do so. This leads to our third prediction.

Prediction 3: In treatment *D&noP* the proportion of players B that choose, in case of delegation, the fair distribution is smaller than the proportion of players A that choose the fair distribution in treatment *noD&noP*.

⁸ Lazear, Malmendier, and Weber (2006) illustrative their experimental design by the situation of a beggar in a street asking for money. Some people give since they derive utility from giving (*like sharing*); others do not give since they do not derive utility from giving (*dislike sharing*). Again others do not derive utility from giving but, in addition, derive disutility from not giving when faced with the request to do so (*dislike not sharing*). The latter will cross the street, i.e. exit the game, to avoid encountering the beggar. Our treatment *D&noP* could be illustrated by the situation in which people can – instead of crossing the street – hand over a joint purse to their companions, i.e. delegate the decision, before passing the beggar.

Notice that Predictions 2 and 3 together imply that the proportion of unfair distributions is higher in treatment *D&noP* than in treatment *noD&noP*. We thus predict that the option to delegate the decision between the two distributions leads to more inequality.

3.2 Predictions concerning the punishment behavior of players C

Consider first treatment *noD&P*. Prior to the punishment stage of the game, there are only two possible moves: player A either chooses the fair or the unfair distribution. Player B cannot take a decision in this treatment. Standard economic theory assumes that all players are purely self-interested and thus predicts that no player C incurs the cost to reduce other players' payoffs – irrespective of player A's preceding decision.

Models of inequity aversion (Fehr and Schmidt 1999, Bolton and Ockenfels 2000) predict that a player C incurs the cost of 1 point to reduce player A's or B's payoffs by 7 points only if the unfair distribution is chosen and if player C's degree of inequity aversion is sufficiently high.⁹ However, these models do not make a prediction about who is punished if the unfair distribution is chosen. Subjects have preferences over outcomes only and it is irrelevant how these outcomes came about. If player C is willing to spend 1 point to reduce other players' payoffs by 7 points, then he is indifferent between reducing either player A's or player B's payoff (or the other player C's payoff, which is higher once the cost of punishing is incurred) as long as these players have higher payoffs.

Models of intention based reciprocity (Rabin 1993, Levine 1998, Charness and Rabin 2002, Dufwenberg and Kirchsteiger 2004, Falk and Fischbacher 2006) also predict that player C punishes only if the unfair distribution is chosen. However, if the unfair distribution is chosen then they unambiguously predict that player A is punished – at least more than player B. The choice of the unfair distribution by player A is an unfriendly move and thus answered in kind. Player B is neutral since he cannot take a decision, and it is not known what he would have chosen in case of delegation. Also the other player C can neither be friendly nor unfriendly towards the punishing player C and should thus not be punished.

Both classes of models thus predict that a proportion of players C punish if player A chooses the unfair distribution. Moreover, while models of inequity aversion are indetermined about who is punished, models of intention based reciprocity make the clear prediction that if there is punishment then player A is punished. This leads to our next prediction.

⁹ In the Fehr-Schmidt model, an inequity averse player C punishes if his $\alpha \geq .75$; see footnote 3.

Prediction 4: In treatment *noD&P* a proportion of players C punish if and only if player A chooses the unfair distribution. If there is punishment, only player A is punished.

Consider now treatment *D&P*. Prior to the punishment stage of the game, there are now four possible scenarios. Either player A chooses the fair or the unfair distribution, or he delegates. If he delegates, then player B either chooses the fair or the unfair distribution.

In models of inequity aversion it does not make a difference whether player A or B takes the decision between the two distributions. If the fair distribution is chosen, this class of models predict that there is no punishment. If the unfair distribution is chosen, they predict that some players C will punish but they do not make a prediction about who is punished.

In models of intention based reciprocity it can be decisive who takes the decision about the fair and the unfair distribution. If player A does not delegate, then the predictions of this class of models in treatment *D&P* do not differ from the predictions in treatment *noD&P*. However, if player A delegates and player B chooses the unfair distribution then players C might interpret both the move by player A and the move by player B as unfriendly. Player B has chosen the unfair distribution, and player A could have chosen the fair distribution but did not to do so. Models of intention based reciprocity then predict that both, player A and player B will be punished. This leads to the next prediction.

Prediction 5: In treatment *D&P*, if player A does not delegate then a proportion of players C punish if and only if player A chooses the unfair distribution. If there is punishment, only player A is punished. If player A delegates and player B chooses the unfair distribution, then a proportion of players C punish. If there is punishment, both player A and player B are punished.

The remaining case is that player A delegates and player B chooses the fair distribution. Player B's decision is then friendly, and he should thus not be punished. However, players C's valuation of player A's decision to delegate should not be affected by the subsequent move of player B. If delegation by player A is interpreted as unfriendly when player B subsequently chooses the unfair distribution, then delegation should also be interpreted as unfriendly when player B subsequently chooses the fair distribution. The extent of the punishment for player A in case of delegation should thus not depend on player B's decision. This leads to our next prediction.

Prediction 6: In treatment *D&P*, if player A delegates and player B chooses the fair distribution, then a proportion of players C punish. If there is punishment, only player A is punished. The extent of the punishment equals the case in which player B chooses the unfair distribution.

3.3 Predictions of choices by players A and B in treatments with punishment

In treatments with punishment we predict that the choice of the unfair distribution is punished by players C. Choosing the unfair distribution is then less attractive, irrespective of whether players A or, in case of delegation, players B must take the decision. We thus predict more fair choices in treatments with punishment.

Prediction 7: The proportion of players A that choose the fair distribution in treatment *noD&P* is higher than in treatment *noD&noP*. The proportion of players B that choose, in case of delegation, the fair distribution in treatment *D&P* is higher than in treatment *D&noP*.

Our experiment is designed to distinguish between different motives for delegation. We predict that even in treatment *D&noP* a proportion of players A delegate the decision between the fair and the unfair distribution to player B. Here, the disutility of being responsible for the unfair distribution constitutes a motive for delegation. In treatment *D&P* there is an additional motive for delegation because unfair choices are punished. We predict that if player A delegates and player B subsequently chooses the unfair distribution, then both player A and player B are punished, whereas solely player A is punished if he does not (or cannot) delegate and chooses the unfair distribution. Since punishment is limited to 7 points, player A will be punished less severely if both players are punished – assuming that all (or most) punishment points are assigned in both cases. Therefore, compared to treatment *D&noP*, in treatment *D&P* there is an additional motive for delegation. Now player A can not only delegate the potential disutility of taking the unfair decision but also – at least in part – the possible punishment by players C. This leads to our final prediction.

Prediction 8: The proportion of players A that delegate in treatment *D&P* is higher than in treatment *D&noP*.

In the next section, we present our results.

4. Results

In this section, we follow the structure of the previous section. We first present the data of the treatments without punishment, then discuss the punishment pattern, and conclude with player A and B's decisions in the treatments with punishment.

In treatment *noD&noP*, only player A has a decision. We find that indeed not all players made the unfair choice: 12 out of 34 players A, which corresponds to 35%, made the fair choice.¹⁰ This confirms our Prediction 1.

In treatment *D&noP*, 29 out of 35 players A did not delegate. Of these players, 23 chose the unfair distribution and 6 chose the fair distribution. There were 6 out of 35 (17%) players A that delegated. This number is significant in the sense that it is significantly different from the 0 delegations in the treatment without delegation (Fisher exact test, $p=0.025$, two-sided). This confirms the first part of Prediction 2, stating that there are some players A that delegate the decision. Figure 3 shows the decisions of players A and B. In this figure, it appears as if the delegations were only made by the players A that would make a fair choice when delegation is not available. However, this conjecture cannot be confirmed statistically (Fisher exact test, $p=0.159$, one-sided¹¹). We compare the distribution of fair and unfair choices in treatment *noD&noP* with the distribution of fair and unfair choices of players A in treatment *D&noP*; that is we do not consider the players A that delegated the decision. Although the 17% of players A who chose the fair distribution when the decision was delegated appears small compared to the 35% of players A who chose the fair distribution when they could not delegate, this difference is far from significant – due to the small number of observations. Finally, we also do not find a significant effect of the delegation option on the outcome (Fisher exact test, $p=0.125$, one-sided). In the treatments without punishment, we find a significant amount of delegation, but we do not find support for more sophisticated hypotheses that the delegation option has an impact on the outcome of the decision. There is one of the 6 players B who chose the fair distribution, the remaining 5 players chose the unfair distribution. We thus cannot confirm Prediction 3.

We now turn to the punishment pattern in the punishment treatments. Figure 2 shows the average punishment points that were assigned to the players A, B and C in the different cases. For instance, the first, black bar shows that in treatment *noD&P* players A that chose

¹⁰ A statistical test for equality with 0% does not make sense, but 35% is for instance highly significantly higher than 10% in a binomial test.

¹¹ We usually report results conservatively, i.e. significance results are reported two-sided and insignificance results are reported one-sided where it makes sense.

the unfair distribution were deducted 3.7 point on average. It is immediately evident that punishment is highest for the person finally causing an unfair outcome – either when player A directly chooses the unfair distribution or when player B does it after being delegated the decision. This is statistically confirmed by comparing the punishment points for the unfair choice with the punishment points for the fair choice. In all these three cases a two-sided Wilcoxon signed rank test is highly significant ($p < 0.01$). By this definition, player B receives no significant punishment in treatment *noD&P*, but players A and B are punished for the unfair outcome in the treatment with delegation also when they are not directly responsible. So for instance player B is punished more when player A chooses the unfair distribution than when player A chooses the fair outcome. When player A delegates and player B chooses the unfair distribution, player B's punishment is much higher than player A's punishment, but the responsibility of player A is somewhat taken into account since his punishment in this case is harsher than player B's punishment when player A directly makes the unfair choice (all these tests are Wilcoxon signed rank tests, $p < 0.01$, two-sided). Finally, there is also very little punishment when the fair outcome is chosen. Also when player A delegates and player B chooses the fair distribution, player A is not punished more than when she chooses the fair distribution herself (Wilcoxon signed rank test, $p = 0.55$).

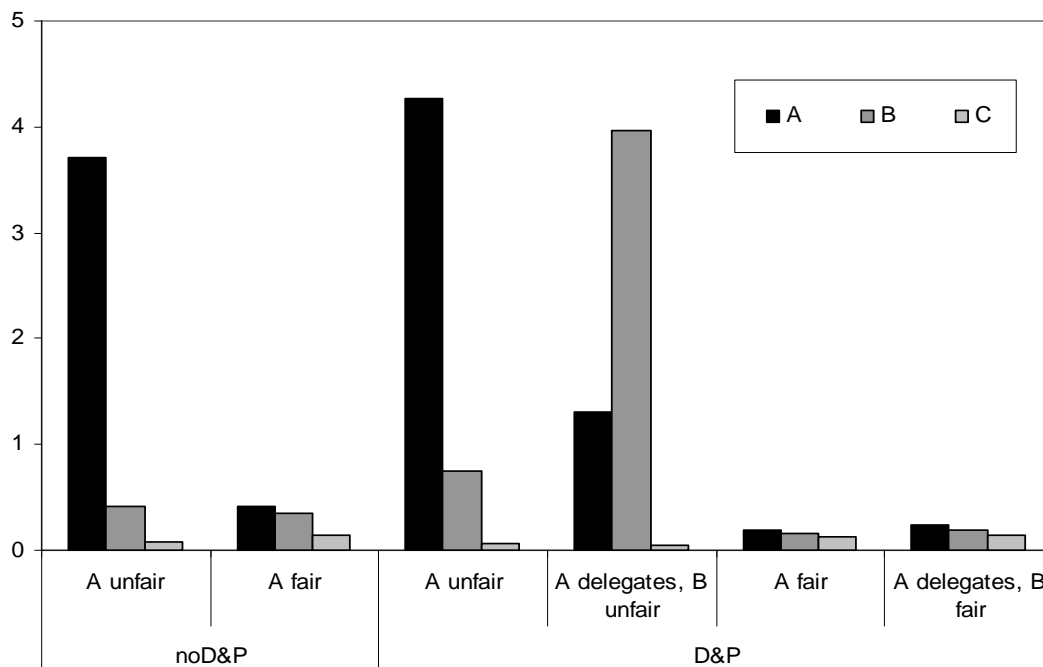


Figure 2: Patterns of Punishment. The bars show the average punishment of the types A, B and C for the different situations and the treatments *noD&P* and *D&P*.

We now turn to the decisions of players A and B in the punishment treatment. All four possible cases are shown in Figure 3. For instance, the second field in the fourth row represents the 22% of observations in treatment *D&P* where player A delegated and player B chose the unfair distribution. The picture shows that the fair distribution is chosen much more often in the treatments with punishment (Fisher exact test, $p < 0.01$). Furthermore, there is a large proportion of delegations. In particular, there are more delegations in the punishment treatment than in the no punishment treatment (Fisher exact test, $p < 0.01$). On the other hand, players B are equally likely to choose the unfair distribution as players A and the unfair distribution results in the two punishment treatments with almost exactly the same probability.

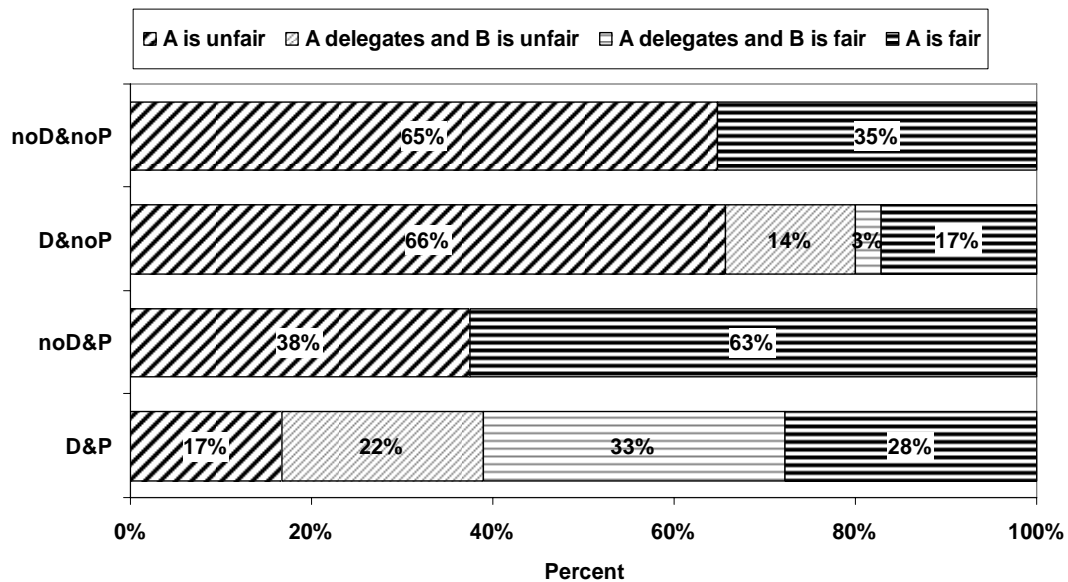


Figure 3: Distributional choices of players A and – in case of delegation – of players B.

To understand these decisions, we study the impact of punishment on player A's income. Figure 4 displays player A's income in the two punishment treatments for all her possible actions. In the no delegation treatment, the unfair choice is statistically more profitable than the fair choice (4.59 vs. 5.30, Wilcoxon signed rank test, $p < 0.01$, two-sided). However, the difference is much smaller than in the treatment without punishment, where the difference equals $9 - 5 = 4$. Thus, it is in line with the more frequent choice of the fair distribution in the treatment with punishment. In treatment *D&P*, there is a clear best strategy for player A. Player A gets a payoff that is more than one point higher if she delegates. This payoff is also significantly higher than the expected payoff of the other two actions (Wilcoxon signed rank test, $p < 0.01$, two-sided). Hence, the comparative static of player A's decisions is fully in line with the observed punishment pattern of the players C.

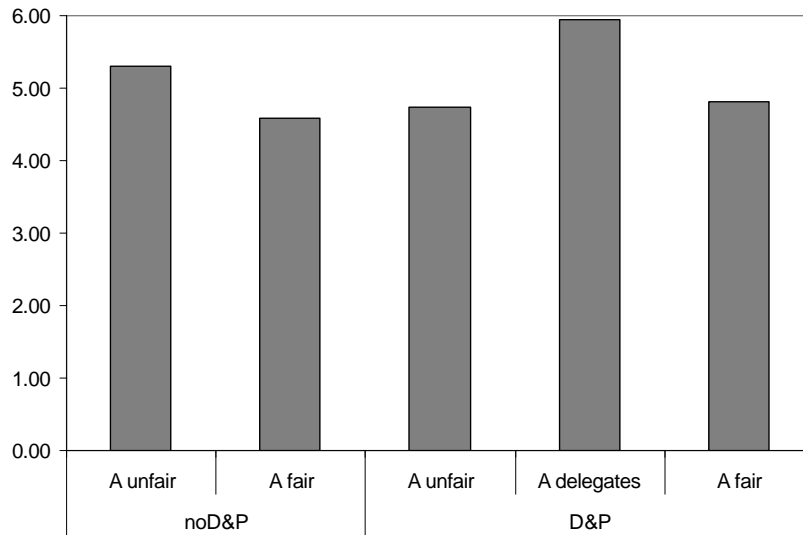


Figure 4: Expected payoff of player A in the punishment treatments.

5. Conclusion

We proposed a new variant of a dictator game that involves delegation and punishment. The experiment was designed to elicit the attribution of responsibility for outcomes of delegated decisions. In our experiment, player A could take a decision between a fair and an unfair distribution of a given endowment. In treatments with delegation, player A could also delegate this decision to player B whose interests were perfectly aligned. Players C were adversely affected whenever the unfair distribution was chosen. In treatments with punishment they could assign costly punishment points and allocate these punishment points at will among the other players. We found a clear pattern of punishment: mainly the player that takes the unfair choice in person is punished while the player that delegated the decision is spared. To distinguish between intrinsic and extrinsic motives for delegation we conducted the treatments without punishment. We find less but significant delegation in these treatments without punishment. This suggests a disutility of taking the unfair decision as an intrinsic motivation for delegation, besides the extrinsic motivation of shifting the blame.

References

- Aghion, Philippe and Jean Tirole, 1997, "Formal vs. Real Authority in Organizations," *Journal of Political Economy*, Vol. 105(1), 1-29
- Andreoni, James, 1989, "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence," *Journal of Political Economy*, Vol. 97, 1447-1458
- Bolton, Gary and Axel Ockenfels, 2000, "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, Vol. 90, 166-193
- Camerer, Colin, 2003, *Behavioral Game Theory*, Princeton, Princeton University Press
- Charness, Gary und Matthew Rabin, 2002, "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics*, Vol. 117, 817-869
- Dana, Jason, Daylian Cain, and Robyn Dawes, 2005, "What you don't know won't hurt me: Costly (but quiet) exit in dictator games," *Carnegie Mellon Behavioral Decision Research Working Paper*
- Dana, Jason, Roberto Weber, and Jason Xi Kuang, 2005, "Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness," *Carnegie Mellon Behavioral Decision Research Working Paper*
- Dufwenberg, Martin und Georg Kirchsteiger, 2004, "A Theory of Sequential Reciprocity," *Games and Economic Behavior*, Vol. 47, 268-298
- Falk, Armin and Urs Fischbacher, 2006, "A Theory of Reciprocity," *Games and Economic Behavior*, Vol. 54 (2), 293-315
- Fehr, Ernst and Klaus Schmidt, 1999, "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, Vol. 114, 817-868
- Fehr, Ernst and Klaus Schmidt, 2006, "The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories", in: *Handbook on the Economics of Giving, Reciprocity and Altruism*, ed. by Serge-Christophe Kolm and Jean Mercier Ythier
- Fischbacher, Urs, forthcoming, "z-Tree: Zurich Toolbox for Ready-made Economic Experiments," *Experimental Economics*
- Greiner, Ben, 2004, "An Online Recruitment System for Economic Experiments", in: *Forschung und wissenschaftliches Rechnen 2003*, ed. by Kurt Kremer and Volker Macho, GWDG Bericht 63, Göttingen, Ges. für Wiss. Datenverarbeitung

- Henderson, Richard, 2006, *Compensation Management in a Knowledge-Based World*, 10th edn., Prentice Hall
- Lazear, Edward, Ulrike Malmendier, and Roberto Weber, 2006, "Sorting in Experiments with Application to Social Preferences," *NBER Working Paper* 12041
- Levine, David, 1998, "Modeling Altruism and Spitefulness in Experiments", *Review of Economic Dynamics*, Vol. 1, 593-622
- Rabin, Matthew, 1993, "Incorporating Fairness into Game Theory and Economics," *American Economic Review*, Vol. 83, 1281-1302