Shifting the Blame:

On Delegation and Responsibility*

Björn Bartling [a)] University of Zurich

Urs Fischbacher [b)] University of Konstanz

July 21, 2008

Abstract:

To fully understand the motives for delegating a decision right, it is important to study responsibility attributions for outcomes of delegated decisions. We conducted an experiment in which subjects were able to delegate the choice between a fair or unfair allocation, and used a punishment option to elicit responsibility attributions. Our results show that, first, responsibility attribution can be effectively shifted and, second, this constitutes a powerful motive for the delegation of a decision right. Furthermore, we propose a formal measure of responsibility and show that this measure outperforms measures based on outcome or intention in predicting punishment behavior.

Keywords: delegation, decision rights, moral responsibility, blame shifting
JEL: C91, D63

* We would like to thank Carlos Alós-Ferrer, Ernst Fehr, Daria Knoch, Michel Maréchal, Klaus Schmidt, Daniel Schunk, Christina Strassmair, Christian Zehnder and seminar participants at the universities of Amsterdam, Konstanz, Malaga, Mannheim, Munich, St. Gallen, Tilburg, and Zurich for helpful comments, as well as Sally Gschwend, Franziska Heusi, and Beatrice John for outstanding research assistance. Support from the Research Priority Program on the "Foundations of Human Social Behavior" at the University of Zurich and the Swiss State Secretariat for Education and Research through the EU-TMR Research Network ENABLE (MRTN CT-2003-505223) is gratefully acknowledged.
[a)] Institute for Empirical Research in Economics, University of Zurich, Bluemlisalpstrasse 10, 8006 Zurich, Switzerland, Tel.: +41-44-634-3722, Fax: +41-44-634-4907, email: bartling@iew.uzh.ch.
[b)] Department of Economics, University of Konstanz, Box 131, 78457 Konstanz, Germany, Tel. +49-7531-88-2652, Fax: +49-7531-88-2145, email: urs.fischbacher@uni-konstanz.de and Thurgau Institute of Economics, Hauptstrasse 90, 8280 Kreuzlingen, Switzerland.

"Princes should delegate to others the enactment of unpopular measures and keep in their own hands the distribution of favours."

Machiavelli[1]

Who is held morally responsible for the outcome of a delegated decision, the person who delegated the decision right or the person who ultimately took the decision? The Stanford Encyclopedia of Philosophy defines moral responsibility in the following way: „To be morally responsible for something, say an action, is to be worthy of a particular kind of reaction — praise, blame, or something akin to these — for having performed it."[2] The question of moral responsibility is thus of economic relevance because praise and blame can constitute effective incentives that affect not only the actual decision but also the decision whether or not to delegate the decision right in the first place. Hence, to fully understand the motives behind the delegation of a decision right, it is important to study the resulting moral responsibility attribution. The aim of this paper is to elicit under controlled laboratory conditions players' responsibility attributions for unfair outcomes in games that allow some players to delegate and others to punish.[3]

The delegation of decision rights is ubiquitous, not only in economic life but also in politics and everyday social interaction. The economic literature has proposed a large number of explanations for why decisions are delegated. The delegate might have lower opportunity costs, be better informed, or equipped with more adequate skills (for an overview of the principal-agent literature see, e.g., Bolton and Dewatripont, 2005). Further explanations include delegation as a

[1] "The Prince" (1961), p. 106.
[2] http://plato.stanford.edu/entries/moral-responsibility/ (retrieved July 21, 2008)
[3] Philosophical reflection on moral responsibility is generally normative, i.e. it poses the question: who ought to be held responsible? We follow a positive, "empirical ethics" approach by studying who is actually held responsible. The normative question is of great importance, also in economics. Indeed, it is at the core of contract theory: contractually stipulated responsibility assigns praise or blame (usually in monetary currency) and thereby provides incentives to take or refrain from certain actions. Since we are solely concerned with moral and not legal or contractual responsibility, in the following we omit "moral" when we refer to responsibility.

commitment device (Schelling, 1960)[4] and incentive provision by delegation (Aghion and Tirole, 1997). The question of responsibility attribution for delegated decisions has caught little attention in the economics literature.[5] In this paper, we provide clear experimental evidence, first, that along with the decision right also the responsibility for the resulting outcome is delegated and, second, that *responsibility shirking*, i.e. *blame shifting*, is a powerful motive for the delegation of a decision right. The results of this paper thus contribute to a better understanding of why decision rights are often delegated.

Consider the following example from the business world. Companies like AlixPartners or Alvarez & Marsal make their living by offering interim management to firms in financial distress. In such cases, a chief restructuring officer (CRO) temporarily replaces or works alongside with the CEO of the troubled firm. In contrast to consultants who act in an advisory capacity, CROs are equipped with extensive decision rights to shepherd, in the most severe cases, an insolvent company through the Chapter 11 process. CROs bring specific expertise and experience, an outside perspective, and supplement incumbent managers in times of intensive work load. Yet these well-established explanations for the delegation of decision rights are not exhausting. The McShane Group, for example, offering "Turnaround Consulting & Crisis Management," advertises its services by stressing that blame can be shifted to CROs: "Moreover, change frequently requires difficult choices and unpopular decisions. The use of an interim executive to move through these decisions and then move on, allows new, permanent leadership to take the helm untainted by any residual negative feelings toward his or her predecessor."[6]

[4] Applications include output and pricing decisions in oligopolistic markets (Vickers, 1985), inflation targeting (Rogoff, 1985), and bargaining (Jones, 1989). Experimental studies of strategic delegation in oligopolistic markets are provided by Huck et al. (2004), in bargaining by, e.g., Schotter et al. (2000) and Fershtman and Gneezy (2001).
[5] An exception is Sliwka (2006) who proposes a career concerns model in which a principal holds one of his agents responsible for a task by announcing his belief that this agent will contribute most to the task. The announced belief can be self-fulfilling as the responsible agent has strong reputational incentives to ensure the success of the task. In contrast, we study players' responsibility attributions in situations where no exogenous ex-ante assignment is made.
[6] http://www.mcshanegroup.com/interim_management.html (retrieved July 21, 2008)

In the political science literature, blame avoidance strategies have been discussed since the 16th century when Machiavelli published his famous book "The Prince," from which the opening quotation is taken. In this tradition, Herring (1940) developed the classic "lightning rod hypothesis." To be effective, he argued, the American president must act "as a generalissimo who devolves upon his generals the responsibility for the attainment of particular objectives. If they fail they can be disgraced and removed; or kicked upstairs to posts of less crucial importance" (p. 112). In modern public choice theory, Fiorina (1982, 1986) applied the concept of blame shifting to regulatory agencies. Under the presumption that actual benefits of regulation can exceed constituents' perceived benefits, he argues that "by charging an agency with the implementation of a general regulatory mandate, legislators […] avoid or at least disguise their responsibility for the consequences of the decisions ultimately made" (1982, p. 47). Another application is international agencies. Vaubel (1986) claims that national politicians "try to get rid of their 'unpleasant' activities, their 'dirty work'" (p. 48). He argues, among other examples, that the International Monetary Fund "relieves its members of unpleasant tasks as well: it imposes policy conditions on borrowing governments which want to evade the responsibility of unpleasant measures; by serving as a bogeyman or scapegoat, it enables the individual lending governments to escape the nationalist resentment which such policy conditions would otherwise create" (p. 49).

Eliciting Responsibility Attribution.—Despite the frequent use and intuitive appeal of the blame shifting motive for delegation, this is to the best of our knowledge the only study to date providing clean evidence on responsibility attribution for delegated decisions. This paper reports data from incentivized choice experiments that are designed to measure responsibility attribution in games allowing for delegation and punishment. We study a variant of a one-shot dictator game in which a first player (the dictator) can decide between an equal (fair) and an unequal (unfair)

allocation of a given endowment. Or, instead of making the decision, he can delegate the decision right to a second player (the delegee) who must then decide between the two allocations. The monetary payoffs of the first and second player are perfectly aligned; both receive a higher monetary payoff if the unfair allocation is chosen. Third players (two receivers) are however adversely affected if the unfair allocation is chosen. They can assign costly punishment points either to the dictator or the delegee or both (or even to the other receiver). Since being responsible means being answerable, i.e. blameworthy or praiseworthy, we take blame as manifested in assigned punishment points as a measure for responsibility attribution. Our results clearly show that responsibility can be effectively shifted. If the dictator delegates the decision right and the delegee makes the unfair choice, then the delegee is harshly punished while the dictator is almost spared. Dictators anticipate this and delegate in the majority of cases. In fact, delegating the decision right was their profit maximizing choice.

Motives for Delegation.—By conducting treatments with and without punishment, the experimental design allows to test whether blame shifting is indeed a motive for the delegation of a decision right. This is strongly confirmed as the share of delegated decisions is three times higher in the treatment with punishment than in the treatment without punishment opportunities by the receivers. Interestingly, a significant share of dictators delegates even in the treatment without punishment. This observation is in line with models assuming that decisions are not only governed by a preference over outcomes but also by a desire to maintain a positive self-image (see, e.g., Konow, 2000; Benabou and Tirole, 2002; Prelec and Bodner, 2003). Delegation offers the chance that the delegee makes the unfair choice, such that the dictator receives the higher payoff without having taken the unfair decision himself, i.e. without having to blame himself.

A Formal Measure of Responsibility.—In the second part of the paper, we investigate more deeply the notion of responsibility, both theoretically and experimentally. We propose a

simple formal measure of a player's responsibility for an outcome of a game. In essence, the measure captures the relative impact of a player's actions on the probability that a certain outcome (e.g. the unfair allocation) results. To test this measure, we conduct two additional treatments that involve delegation to a random device and asymmetric action spaces. An econometric comparison of different punishment motives shows that the responsibility measure can explain more of the variation in the punishment patterns in the different treatments than a purely outcome based measure (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) or a measure of intention based reciprocity (Rabin, 1993; Levine, 1998; Charness and Rabin, 2002; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006). This final result lends support to our interpretation that assigned punishment points reflect responsibility attribution.

The paper is organized as follows. Section I explains our experimental design in detail. Section II presents the results on punishment patterns, frequency of delegation, and allocation choices. Section III proposes a formal measure of a player's responsibility for an outcome of a game. Section IV analyzes additional treatments to test the responsibility measure. Section V provides an econometric comparison of different punishment motives. Section VI concludes.

I. Experimental Design

We implemented a variant of a dictator game that allows for delegation and punishment. There are groups of four players. Each group consists of one player A (the dictator), one player B (the potential delegee), and two player Cs (the receivers). Player A or B can decide between an equal and an unequal allocation of 20 points among the four players in the group. The equal (fair) allocation assigns 5 points to each of the four players; the unequal (unfair) allocation assigns 9 points each to players A and B and 1 point each to both player Cs.

We considered treatment variations along two dimensions: with or without delegation and with or without punishment. All treatments were played one-shot. Figure 1 gives an overview.

	no punishment	punishment
no delegation	*noD&noP*	*noD&P*
delegation	*D&noP*	*D&P*

FIGURE 1. EXPERIMENTAL DESIGN

In treatments without delegation, player A must decide between the fair and the unfair allocation. Player B cannot take any decision. In treatments with delegation, player A can, instead of making the decision himself, delegate the decision right to player B. If player A delegates, then player B must make the decision. He cannot refuse to make the decision nor delegate it to yet another player. If player A does not delegate, then player B cannot make any decision.

In treatments without punishment, the player Cs cannot make decisions. In treatments with punishment, one player C is randomly selected. The selected player C can — given player A's and, in case of delegation, player B's decisions — assign costly punishment points to players A, B, and the other player C. He can spend one of his points to reduce the other players' points by altogether 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 to all three other players. But no player can be taken more points than he received in the chosen allocation. The selected player C can also decide to assign less than seven punishment points and leave the unassigned points void.[7] The player C that is not selected cannot take any decision. Figure 2 summarizes the sequence of possible moves in the game.

[7] We are mainly interested in relative punishment and not in punishment levels. This allows employing our very simple punishment technology that focuses—given a player C wants to punish—on the assingment of a given number of punishment points to different players rather than on the question how much to punish in total (even though a punishing player C has the option not to assign all seven points).

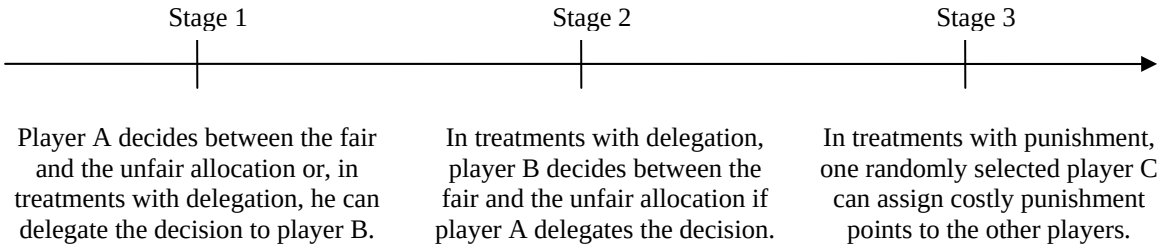

FIGURE 2. SEQUENCE OF POSSIBLE MOVES IN OUR GAMES

We applied the strategy method for player Cs. Each player C had to decide (in a randomized order) how many punishment points to allocate to players A, B, and C in all possible situations before they knew how player A and — in case of delegation — player B decided and whether they were selected to be able to punish or not. In the treatment without delegation there are two situations: player A is either fair or unfair. In the treatment with delegation there are four situations: player A does not delegate and is either fair or unfair, or player A delegates and player B is either fair or unfair. All situations occurred with a probability of at least 16 percent, so there were ample incentives in all situations to make a considerate decision. We did not use the strategy method for player Bs; they made a decision only when the decision right was in fact delegated.

In our games, we had two player Cs for efficiency reasons. We are mainly interested in the responsibility attribution, i.e. in the punishment patterns that arise in the different situations. With only one player C in a group we elicit this information only from a third of our subjects; with two player Cs we obtain punishment decisions from half of our subjects. Moreover, if we had only one player C, player As and Bs would each gain only half of what player C loses, which might result in most player As and Bs choosing the fair allocation; e.g. (7, 7, 1) vs. (5, 5, 5). If, alternatively, player As and Bs each obtained what the single player C looses, we would introduce efficiency as a confounding factor; e.g. (9, 9, 1) vs. (5, 5, 5).

Procedural Details.—The experiment was computerized with the software "z-Tree" (Fischbacher, 2007). The recruitment was conducted with the software "ORSEE" (Greiner, 2004). Subjects were students from the University of Zurich and the Swiss Federal Institute of Technology in Zurich. Economics and psychology students were not eligible to participate. All sessions took place at the Institute for Empirical Research in Economics at the University of Zurich. 144, 128, 140, and 136 subjects participated in the treatments *D&P*, *noD&P*, *D&noP*, and *noD&noP*, respectively. Another 132 and 144 subjects participated in the control treatments *random* and *asymmetry* (see Section IV), respectively. No subject participated more than once.

Subjects were randomly assigned a role as player A, B, or C upon arrival at the lab. They received written instructions including comprehension questions that had to be answered correctly. A summary of the instructions was read aloud to ensure common knowledge of the game. All treatments were framed in a neutral manner.[8]

Sessions without punishment lasted for about 45 minutes; sessions with punishment for about 60 minutes. Each experimental point was converted into CHF 3 (about $2.4 at that time) at the end of the experiment. On average, subjects earned CHF 25 (about $20) in the sessions without punishment and CHF 22.90 (about $18.40) in the sessions with punishment, which includes a show-up fee of CHF 10 (about $8). We also conducted an incentivized belief elicitation session with 32 subjects. This session lasted for about 60 minutes and subjects earned CHF 23.10 (about $18.60) on average.

[8] Participants were called player A, B, and C. The two allocations were called allocations 1 and 2; we did not use value laden terms like "fair" and "unfair." Player Cs could "reduce" other players' points; we did not write "punish."

II. Results

A. *Punishment Patterns*

The punishment treatments *noD&P* and *D&P* allow eliciting player Cs' responsibility attributions for the unfair outcome in our games. Since being responsible for an unfair outcome means being blameworthy for it, we measure responsibility attribution by measuring assigned punishment points. We find a clear pattern of punishment. When the fair outcome results, then there is almost no punishment. When the unfair outcome results, then mainly the player that chooses the unfair allocation is punished, while the other players are almost spared. Most importantly, this pattern holds even if player A delegates and player B subsequently chooses the unfair allocation. Thus, assigned punishment points, i.e. responsibility attribution can be effectively shifted by delegating the decision right.

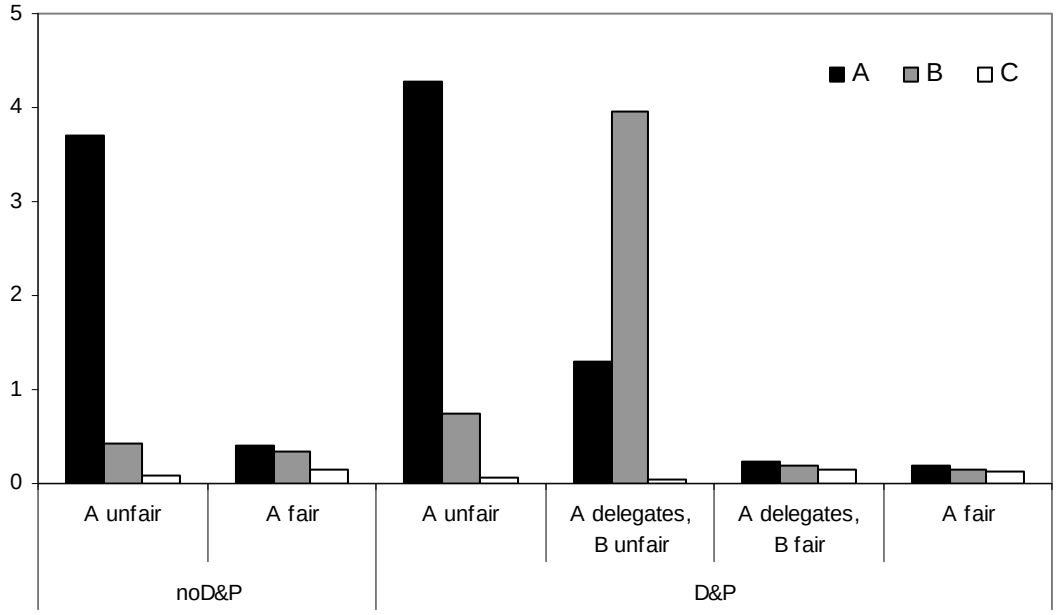

FIGURE 3. AVERAGE PUNISHMENT LEVELS IN THE DIFFERENT SITUATIONS

Figure 3 shows the average punishment points that were assigned to players A, B, and the respective other player C in the different situations. The exact values can be read from Table 1 in

Section III. For instance, the left black bar shows that in treatment *noD&P* player A was deducted 3.70 points on average if he chose the unfair allocation. From the figure it is immediately evident that the average punishment for the player with the decision right is higher if he chooses the unfair allocation than if he chooses the fair allocation. This is statistically confirmed by comparing the respective punishment points for the fair and the unfair allocation choice. In all three comparisons, two-sided Wilcoxon signed rank tests are highly significant ($p<0.01$). Moreover, in all three situations in which the unfair allocation is chosen, average punishment is highest for the player that made the allocation choice. This is statistically confirmed by comparing the punishment points for that player with the punishment points for the respective other two players. In all three situations, two-sided Wilcoxon signed rank tests are highly significant ($p<0.01$).

The important finding is that in treatment *D&P*, player A receives an average punishment of 4.27 points if he chooses the unfair allocation, but a much lower average punishment of only 1.31 points if he delegates and player B subsequently chooses the unfair allocation. Inversely, player B receives an average punishment of only 0.75 points if player A chooses the unfair allocation, but a much higher average punishment of 3.96 points if he chooses the unfair allocation after delegation by player A. In both comparisons, two-sided Wilcoxon signed rank tests are highly significant ($p<0.01$). This result shows that by delegating the decision right to player B, player A also delegates most of the punishment for the unfair outcome to player B.

This finding is further illustrated in Figure 4. The figure shows the individual player Cs' assignments of punishment points to players A and B in treatment *D&P*. The left panel shows the situation in which player A chooses the unfair allocation. The right panel shows the situation in which player B chooses the unfair allocation after delegation by player A. Grey circles above (below) the 45-degree line indicate player Cs who punish player A more (less) than player B. For

instance, the top left circle in the left panel shows that in the situation in which player A chooses the unfair allocation, 29 (out of 71)[9] player Cs assign all seven punishment points to player A and no punishment points to player B; the circle on the origin indicates that 19 player Cs did not punish at all. Figure 4 clearly shows that if player A makes the unfair allocation choice, then of those player Cs who punish, almost all punish player A more than player B. In contrast, if player A delegates the decision right and player B makes the unfair allocation choice, then of those player Cs who punish, the vast majority punishes player B more than player A.

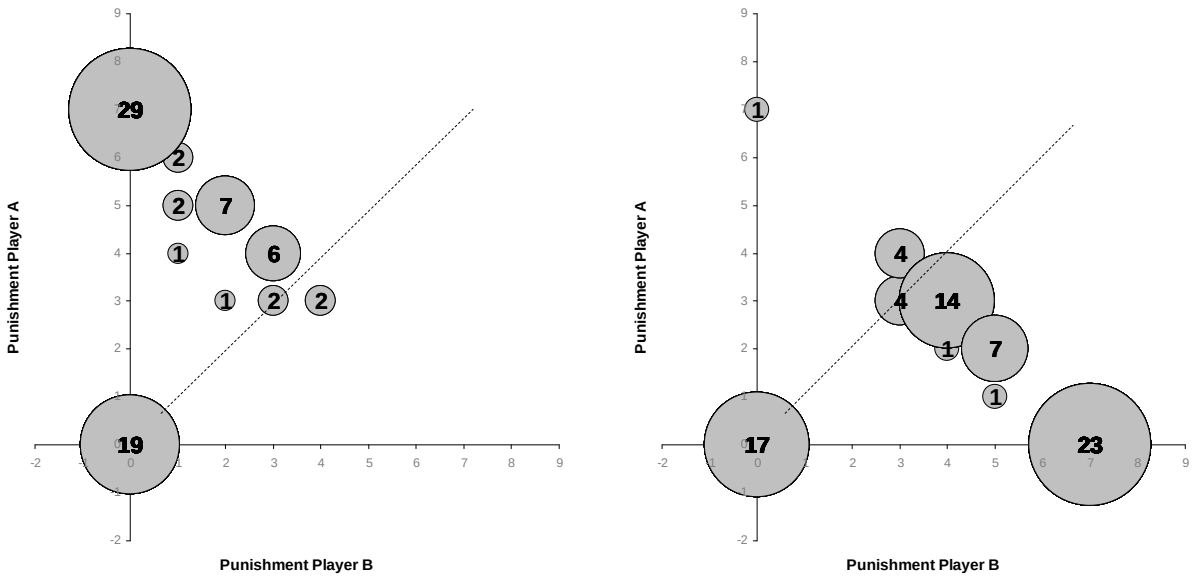

FIGURE 4. INDIVIDUAL PLAYER CS' ASSIGNMENTS OF PUNISHMENT POINTS

Notes: The left panel shows the situation in which player A chooses the unfair allocation in treatment *D&P*. The right panel shows the situation in which player A delegates and player B chooses the unfair allocation. Numbers in circles indicate the number of observations.

Furthermore, from Figure 3 it can be seen that in treatment *D&P* players A and B receive more punishment if the respective other player makes the unfair than if he makes the fair allocation choice. We find that player B is punished more if player A chooses the unfair

[9] Due to a programming error in the sequencing of the different situations, the punishment points of one player C in the two situations in which (i) player A chooses the unfair allocation and (ii) player A delegates and player B chooses the fair allocation were not recorded in our first *D&P* session. As a result, we have only 71 observations in the left panel of Figure 4, while we have all 72 observations in the right panel.

allocation (0.75 points) than if player A chooses the fair allocation (0.15 points). Also player A is punished more if player B chooses the unfair allocation (1.31 points) than if player B chooses the fair allocation (0.24 points). In both comparisons, two-sided Wilcoxon signed rank tests are highly significant ($p<0.01$). Importantly, player A is punished significantly more if he delegates and player B chooses the unfair allocation (1.31 points) than player B if player A chooses the unfair allocation (0.75 points); two-sided Wilcoxon signed rank test ($p<0.01$). This observation can also be made in Figure 4. There are more player Cs in the right panel who assign a larger share of total punishment to player A (i.e. player Cs that are located close to the 45-degree line), than there are player Cs in the left panel who assign a larger share of total punishment to player B. This finding shows that there are limits to punishment shifting (which we will further address in our control treatments in Section IV). After all, if player B chooses the unfair allocation, player A could have chosen the fair allocation but instead decided to delegate. Player B, in contrast, did not have the opportunity to secure the fair allocation (hence did not neglect that opportunity) if player A chooses the unfair allocation directly.[10]

Theoretical Predictions.—What do different models of social preferences predict about the punishment patterns? Outcome based models of inequity aversion (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) predict that sufficiently inequity averse player Cs incur costs to reduce inequality. These models thus correctly predict that there is some amount of punishment if the unfair allocation is chosen and no punishment if the fair allocation is chosen. But these models do not predict who is punished, because to reduce inequality it does not matter whether player A, player B, or both are punished.

[10] In treatment *noD&P*, there is not even the option to delegate the decision right to player B, which seems to further separate player B from player A's decision. In this treatment, player B is not punished significantly more if player A chooses the unfair allocation (0.42 points) than if player A chooses the fair allocation (0.34 points); one-sided Wilcoxon signed rank test, $p=0.37$.

Models of intention based reciprocity (Rabin, 1993; Levine, 1998; Charness and Rabin, 2002; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006) predict that reciprocal player Cs respond to unkind behaviors by player A and/or player B by assigning punishment points. If player A chooses the unfair allocation — a clearly unkind action — then intention based models predict that player A will be punished. If player A delegates and player B chooses the unfair allocation, then player Cs might interpret both actions as unkind. Player B chose the unfair allocation and player A could have chosen the fair allocation but instead delegated the decision right. Models of intention based reciprocity thus correctly predict that player B and, to some extent, also player A will be punished.[11] But since player C's belief about the intention behind player A's decision to delegate cannot depend on player B's subsequent action, player A's punishment should not depend on player B's subsequent action.[12] This is, however, not what we observe: if player A delegates, then he is significantly more punished if player B chooses the unfair allocation than if player B chooses the fair allocation (see above). Indeed, if player B chooses the fair allocation after delegation by player A, then player A is not punished more than if player A chooses the fair allocation directly (one-sided Wilcoxon signed rank test, $p=0.65$).

In this paper, we interpret assigned punishment points as attributions of responsibility for the unfair outcome of the game. In Section III, we propose a simple formal measure of responsibility. This measure allows to derive exact predictions about the punishment patterns

[11] It is sensible to assume that delegation is less kind than choosing the fair allocation but less unkind than choosing the unfair allocation. In the next section, we discuss a measure of unkindness that confirms this intuition.

[12] If player C's belief about player A's belief about player B's choice depends on player B's actual choice, then the following equilibriums can be constructed. First, player C does not punish player A if player B is fair, because if player B chooses the fair allocation, then player C believes that player A believed that player B would be fair. And second, player C does punish player A if player B is unfair, because if player B chooses the unfair allocation, then player C believes that player A believed that player B would be unfair. For these equilibriums to be convincing, one has to assume that player C thinks that player A is better informed about the distribution of fair player Bs, because player C could then update his belief about player A's belief about player B by observing player B's actual choice. However, since the role assignment in our experiment was random, player C cannot rationally believe that player A holds superior information about player B.

based upon the notion of responsibility and to compare its explanatory power with other punishment motives like outcome or intention.

B. *Frequency of Delegation and Players' Allocation Choices*

Comparing the shares of delegated decisions in treatments *D&noP* and *D&P* allows making inferences about the motives behind the delegation of a decision right. We find that the share of player As that delegate the decision right is more than three times higher in treatment *D&P* than in treatment *D&noP*. Thus blame shifting, i.e. responsibility shirking, can be an important motive for the delegation of a decision right.

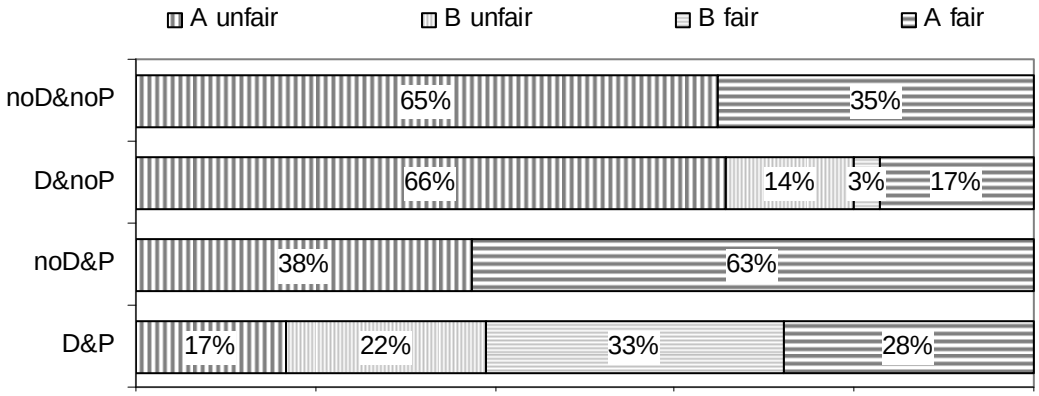

FIGURE 5. DELEGATION AND ALLOCATION CHOICES

Figure 5 shows the share of delegated decisions and player As' and Bs' allocation choices. For instance, the left part of the top bar shows that in treatment *noD&noP*, 65% of player As chose the unfair allocation. A first inspection of Figure 5 shows that our treatments replicate two well established experimental results. First, some people exhibit fair behavior in dictator games and, second, the threat of punishment increases the share of fair choices because punishment can render it optimal also for selfish people to conform to fairness norms (e.g., Forsythe et al., 1994). In treatment *noD&noP*, 35% of player As (12 out of 34) chose the fair allocation. In treatment *noD&P*, the share of fair player As rises to 63% (20 out of 32), which is significantly higher

(two-sided Fisher exact test, p=0.048). The same pattern holds in the treatments with delegation. In treatment *D&noP*, 20% of the players (7 out of 35) chose the fair allocation. In treatment *D&P*, the share of fair players rises to 61% (22 out of 36), which is again significantly higher (two-sided Fisher exact test, p<0.01).

The important finding is that the share of delegated decision is more than three times higher in the treatment with than in the treatment without punishment. In treatment *D&noP*, 17% of player As (6 out of 35) delegated the decision right. In treatment *D&P*, the share of delegated decision rises to 56% (20 out of 36), which is significantly higher (Fisher exact test, p<0.01).[13] This finding demonstrates the strong effect of blame shifting on the decision whether or not to delegate. Notice that our design abstracts from other reasons for delegation such as informational asymmetries, skill, opportunity costs, or commitment. The virtue of a controlled laboratory experiment is that we can abstract from these (potentially important) motives: players A and B are designed completely symmetric — except that player A holds the decision right initially and can decide whether or not to delegate it.

Looking at expected payoffs, we find that in treatment *D&P* delegation was highly significantly more profitable than making either allocation choice oneself (Wilcoxon signed rank tests, p<0.01, two-sided).[14] Delegation led to an expected payoff of 5.93, while the choice of the fair and unfair allocations led to expected payoffs of 4.80 and 4.73, respectively. Consistent with this, the majority of player As delegated. The payoff maximizing choice in treatment *D&noP* was choosing the unfair allocation because not all player Bs made the unfair choice after delegation

[13] Of the six player Bs that could take a decision in treatment *D&noP*, only one chose the fair allocation. Of the 20 player Bs that could take a decision in treatment *D&P*, 12 chose the fair allocations. Fisher exact tests show that player Bs' choices are not significantly different from the allocation choices by the player As that did not delegate. Due to the small number of observations, however, insignificance is to be expected.

[14] We calculate expected payoffs given that 40 percent of player Bs are unfair and given the respective punishment patterns of player Cs. A robustness check shows that the test remains highly significant as long as at least 23% of player Bs are unfair.

(i.e. delegation is at least weakly dominated). Consistent with this, the majority of player As chose the unfair allocation.

Interestingly, a significant share of dictators delegated the decision right even in treatment *D&noP*, where avoiding punishment cannot be a reason for delegation.[15] Choosing the unfair allocation might however involve psychological costs, which could be avoided by delegating. This interpretation conforms to economic models in which decisions are not only governed by a preference over outcomes but also by the desire to avoid cognitive dissonance or to maintain a positive self-image (see, e.g., Konow, 2000; Benabou and Tirole, 2002; Prelec and Bodner, 2003). It is in player A's self-interest to choose the unfair allocation that yields the higher monetary payoff. But making the unfair choice might be in conflict with his resolution to divide fairly, or he might reveal (to himself) that he is a selfish and greedy person. If instead player B chooses the unfair allocation, then player A receives the higher payoff without experiencing cognitive dissonance between self-interest and fairness, or without sending a signal (to himself) that he is a person who increases his payoff at the expense of someone else.[16]

III. A Measure of Responsibility

In this paper, we are interested in player Cs' responsibility attributions for the unfair outcome in our games. Empirically, we measure responsibility attribution by measuring assigned punishment points. In this section, we propose a formal measure of responsibility. The measure is not meant to be comprehensive of the complex meaning of the notion of responsibility.[17] It is rather meant to be simple but nevertheless to capture a basic understanding of what it means to be responsible

[15] Assuming common knowledge of rationality and selfishness, player A would be indifferent between choosing the unfair allocation and delegating to player B. Yet by adding a tiny amount of uncertainty about player B's subsequent action, player A would strictly prefer being unfair himself, i.e. only being unfair oneself is trembling hand perfect.
[16] For related experimental findings see Dana et al. (2006, 2007), Hamman et al. (2007), and Lazear et al. (2006).
[17] See, for example, the entry on "moral responsibility" in the Stanford Encyclopedia of Philosophy http://plato.stanford.edu/entries/moral-responsibility/ and the references given therein.

for an event, i.e. for an outcome of a game. The reason to propose a formal measure of responsibility is twofold. First, the measure allows deriving exact predictions about player Cs' assignment of punishment points in the different situations. Second, if the measure is successful in explaining the punishment pattern (relative to other measures like outcomes or intentions that capture punishment motives), then this finding would support our interpretation that assigned punishment points reflect responsibility attributions.

How can a player's responsibility for an outcome of a game (e.g. the unfair allocation results) be captured? Our measure assigns most responsibility to the person who affected the outcome most. More precisely, it calculates the impact of a player's action(s) on the probability that a certain outcome results. An observer who evaluates a player's responsibility holds a belief about all players' strategies in the game. This belief determines at each node in the game the probability that a certain outcome results. A player's actual action at a node can change this probability (e.g. if the observer's belief is a non-degenerate probability distribution over the set of actions at that node). According to our measure, a player takes on responsibility for an outcome (e.g. the unfair allocation in our games) if and only if his actions increase the probability that this outcome will result. If more than one player increases this probability, then each player's share in the overall probability increase is calculated. In the calculation of the overall probability increase, we do not include moves of nature. If, for example, a single player's action and a move by nature both increase the probability for a certain outcome, then the single player is fully responsible. A player's responsibility is thus not diluted by moves of nature. This captures the idea that only people but not "chance" can be responsible for an outcome. Finally, if the outcome of interest does not realize, then our measure of responsibility is zero. This captures the idea that nobody must be held responsible for an event that did not happen. In the following, we give a formal, more detailed explanation of the responsibility measure.

Consider a multi-player extensive form game with complete and perfect information and a finite number of stages. Let $\Im = \{1,...,I\}$ be the set of players and i be a player in the game. N denotes the set of nodes and N_i the set of nodes where player i has the move. Let $m,n \in N$ be nodes of the game. If node n follows node m (directly or indirectly), we denote this by $m \to n$. Let $\nu(m,n)$ be the unique node that directly follows node m on the path from m to n. Let F be the set of end nodes and f a single end node. The payoff function for player i be defined as $\pi_i : F \to \Re$, where $\Re$ is the set of real numbers. Let A_n be the set of actions in node n. Let $P(A_n)$ be the set of probability distributions over the set of actions in node n. $S_i = \Pi_{n \in N_i} P(A_n)$ is player i's strategy space, $s_i \in S_i$ a strategy of player i, and $S = \Pi_{i \in \Im} S_i$ the strategy space of the game. Let β denote a player's belief about all players' strategies, i.e. a probability distribution over S, and $g(\beta)$ the probability density of a distribution of beliefs. Finally, assume an outcome function $\omega : F \to \Re$. In our games, we are interested in a player's responsibility for the unfair allocation, so the outcome function is an indicator function that equals 1 if the unfair allocation results and 0 if the fair allocation results.[18] Let $\overline{\omega}(n|\beta)$ denote the expected value of ω at node n given belief β. In our games, $\overline{\omega}(n|\beta)$ thus measures the probability of reaching the unfair allocation.

We now define the responsibility of a player i for an outcome ω of a game in three steps. First, player i's *raw responsibility* for outcome ω in node n, given the belief β, is defined as

$$r_i^0(n,\omega|\beta) = \max\left\{\sum_{\substack{m \in N_i \\ m \to n}} \left(\overline{\omega}(\nu(m,n)|\beta) - \overline{\omega}(m|\beta)\right), 0\right\}, \quad (1)$$

[18] We use this outcome function instead of the payoff function of player C to simplify the presentation. The payoff of player C equals 5 or 4 if the fair allocation is chosen, and 1 or 0 if the unfair allocation is chosen. Therefore, the main variance in the payoff of player C is caused by the outcome, i.e. whether the fair or the unfair allocation is chosen.

the sum of changes in $\overline{\omega}$ that resulted from a player's moves along the path to n. If player i's moves resulted in a net decrease of the probability that outcome ω will realize, then his raw responsibility is defined to be zero. In our games, if a player reduces the probability that the unfair allocation will realize, he is not responsible should it finally realize.

Second, player i's *share in total raw responsibility* for the outcome ω in node n, given belief β, is defined as

$$r_i(n,\omega|\beta) = \frac{r_i^0(n,\omega|\beta)}{\sum_{j\in\Im} r_j^0(n,\omega|\beta)}, \tag{2}$$

which measures a player's share in the probability changes that lead to outcome ω. A player's share lies between 0 and 1. It is important to notice that the denominator sums over all *players'* raw responsibilities, but that the summation does not include *nature*. This is to capture that responsibility can only be borne by people but not by "chance."

Finally, player i's *responsibility* for outcome ω in node n, given a probability distribution $g(\beta)$ over beliefs, is defined as

$$R_i(n) = \int \overline{\omega}(n|\beta)\ r_i(n,\omega|\beta)\ g(\beta)d\beta\,. \tag{3}$$

There are two features to notice in the definition of $R_i(n)$. First, player i's share in total raw responsibility is weighted by $\overline{\omega}(n|\beta)$, the probability at node n that outcome ω will result, given belief β. In our games, if n is a node where player C makes his punishment decisions, then irrespective of β this probability is either 1 or 0, i.e. the event did either happen or not. Thus, if the unfair outcome is not chosen, then a player's responsibility is defined to be zero — even if his action(s) increased the probability of the unfair outcome. Second, the responsibility measure allows for belief uncertainty, i.e. it does not require that an observer holds a point belief about the players' strategies but allows for a non-degenerate distribution over beliefs. A player's

responsibility is then given by integrating over this distribution. Alternatively, the integral can be interpreted as averaging over heterogeneous beliefs of multiple observers whose responsibility attribution is to be captured.[19]

A Measure of Intentions.—How is the notion of responsibility different from a notion of intentions? Define a player's intention, or *un*kindness, behind a move at node m as

$$\varphi(m,\omega|\beta)=\frac{\bar{\omega}(n|\beta)-\bar{\omega}(n^{\min}|\beta)}{\bar{\omega}(n^{\max}|\beta)-\bar{\omega}(n^{\min}|\beta)}, \tag{4}$$

where n is the node that the action taken at node m leads to. Out of all nodes that can be reached from m, $n^{\max}$ is the node that maximizes $\bar{\omega}$ and $n^{\min}$ is the node that minimizes $\bar{\omega}$, given belief β. In our games, the "best" intention (the least unkind action) is thus given by a value of 0, the "worst" intention by a value of 1. If $\bar{\omega}(n^{\max}|\beta)-\bar{\omega}(n^{\min}|\beta)=0$, then $\varphi(m,\omega|\beta)\equiv 0$, i.e. a player who cannot influence the probability of outcome ω is attributed an intention of zero.

There are several important differences between the notions of intention and responsibility. First, the measure of intentions does not depend on the outcome that is finally realized. An action can be unkind even if the intended outcome does not realize (e.g. for reasons beyond the player's control). In contrast, if outcome ω does not realize, then the measure of responsibility is zero, because nothing happened that someone must be held responsible for. Second, while the measure of intentions is calculated by evaluating a player's action relative to his action space (i.e. relative to what he could have done), the measure of responsibility is calculated by determining the impact of a player's action on the probability that outcome ω will realize. The action space is relevant for the responsibility measure only inasmuch as the observer's belief puts weight on the possible actions. That is, by adding a very unkind action possibility (a new $n^{\max}$) that is however believed to be never chosen, the responsibility measure

[19] Indeed, this is how we calculate the players' responsibility measures in Table 1 below.

would remain unchanged while the intentions measure would improve, because a given action appears friendlier relative to this new, very unkind action possibility. Third, the measure of intentions is not sensitive to whether a *person* or *nature* is going to make a move subsequent to a player's action. In the context of our games, suppose the belief about player B's move equals the commonly known probability with which a dice randomizes between the fair and the unfair allocation. For the intention measure it does not make a difference whether player A delegates to the dice or to player B. The reason is that the probability for the unfair outcome is the same in both cases. For the responsibility measure, however, it makes a crucial difference, because only moves by people (and not chance) enter the determination of a player's share in total raw responsibility. Finally, while multiple players in a game can be attributed an intention measure of one, i.e. the unkindest intention, the sum of different players' responsibility measures cannot exceed one. This captures the idea that it is not possible to have multiple players, each of which is fully responsible for an outcome.[20]

Belief Elicitation.—Since both the responsibility and intention measure depend on beliefs, we conducted a belief elicitation session. We successively explained the participants our treatments *noD&P* and *D&P* and told them that we had conducted these experiments in our laboratory in the past. Control questions had to be answered to check the understanding of the instructions. The subjects then had to guess the frequencies with which the player As and — in case of delegation — the player Bs choose the fair and unfair allocations. The answers were

[20] In the economics literature, the term "responsibility" is used with diverse meanings. For example, Charness and Jackson (2007) analyze the role of responsibility in strategic risk-taking. They define a player to have responsibility if his action determines the payoff also of another, passive player. Prendergast (1995) proposes a theory of responsibility in organizations and defines responsibility as "span of managerial control" (p. 388). Manove (1997) analyzes how pay and promotion prospects should be related to job responsibility, which he defines as "the variation in the value of job outcomes over the feasible range of worker effort" (p. 86). These papers have in common that they define responsibility ex-ante: responsibility is determined by a player's action set or scope of his actions. In contrast, we are interested in ex-post responsibility, i.e. in assigning responsibility for an outcome or event that is the result of one or many players' actions.

incentivized as the participants earned more experimental points the better their beliefs matched actual play. The numbers in Table 1 are calculated based on this belief elicitation session.[21]

TABLE 1—ASSIGNED PUNISHMENT POINTS AND PUNISHMENT MOTIVES

		noD&P		*D&P*			
		A unfair	A fair	A unfair	B unfair	B fair	A fair
average punishment	A	3.70	0.41	4.27	1.31	0.24	0.19
	B	0.42	0.34	0.75	3.96	0.20	0.15
responsibility	A	1	0	1	0.02	0	0
	B	0	0	0	0.98	0	0
intention	A	1	0	1	0.34	0.34	0
	B	0	0	0	1	0	0
outcome	A	1	0	1	1	0	0
	B	1	0	1	1	0	0

Table 1 shows the average punishment in the different situations in treatments *noD&P* and *D&P* together with the predictions of the punishment motives responsibility, intention, and outcome. For example, the left column shows that in treatment *noD&P* in the situation in which player A chooses the unfair allocation, player As are punished by 3.70 points and player Bs by 0.42 points on average. In the rows underneath, the intensity of the three punishment motives for players A and B are shown separately. The responsibility and the intention measure both assign the highest value of 1 to player A and the lowest value of 0 to player B. Player A is fully responsible, because the total probability increase of the unfair allocation is caused by player A. Player B is not responsible at all, because he cannot make a decision, hence he cannot influence

[21] The numbers are derived by averaging over the values of the responsibility measure that result from each participant's belief in the belief elicitation session. The interesting values are contained in the column "B unfair." The very low value of 0.02 for player A's responsibility stems from the fact that the participants believe on average that player As and Bs choose very similarly between the two allocations, with player Bs being only slightly less fair. Therefore, by delegating to player B, the probability that the unfair allocation will result is increased only very little. The value of 0.34 for player A's intention (i.e. unkindness) behind delegating to player B (for both cases, "B unfair" and "B fair"), results from the participants' average belief that 34% of player Bs choose the unfair allocation. This is a reasonably accurate belief as the actual value was 40%. To assess the robustness of the responsibility measure, we also calculated values based on a uniform belief distribution and find that the order of the values is preserved.

the probability of the unfair allocation. Also the measure of intention assigns a value of 1 to player A, because choosing the unfair allocation is the unkindest action that he can take. Since player B cannot make a decision, he cannot act in an unkind manner and is thus assigned a value of 0. Purely outcome based models predict that if the unfair allocation is chosen, player Cs punish to reduce the inequality of the outcome. Since both player A and player B receive a higher payoff than the player Cs, both are assigned a value of 1.

Table 1 shows that that notion of responsibility captures the punishment behavior in the treatments *noD&P* and *D&P* rather well: whenever it predicts a different level of responsibility in two situations then average punishment in the two situations differs in the predicted direction. The purely outcome based approach makes the wrong prediction that punishment in case of the unfair allocation does not depend on who took the decision. The intention based approach makes the wrong prediction that punishment does not depend on the outcome after delegation by player A (see footnote 13). In the next section, we propose two control treatments to test the measure of responsibility and to further disentangle the notions of intention based reciprocity and responsibility. In Section V, we provide an econometric comparison of different punishment motives.

IV. Control Treatments

This section reports on results in two control treatments that we conducted to test the measure of responsibility and to further distinguish between the punishment motives responsibility and intention. Both control treatments are variants of treatment *D&P*. The first control treatment ("*random*") differs from treatment *D&P* in that player A can now delegate the allocation choice to a dice but not to player B. The dice chooses the fair allocation with probability 0.6, which exactly matches the 12 out of 20 player Bs who made the fair choice in treatment *D&P*. This

randomization probability was common knowledge.[22] Player B cannot make a choice in treatment *random*. The second control treatment ("*asymmetric*") differs from treatment *D&P* in that player A can only choose the fair allocation or delegate to player B, i.e. he cannot choose the unfair allocation himself. As in treatment *D&P*, player B can decide between the fair and the unfair allocation if player A delegates. In both control treatments, the punishment possibilities by player Cs were exactly as in treatment *D&P*.

According to our responsibility measure, only people but not chance (e.g., a dice or a computer) can take on responsibility. In treatment *random*, we thus predict that in case the unfair allocation results after delegation, player A will be punished more for delegating to the dice than for delegating to player B. In contrast, for a model of intention based reciprocity it does not make a difference, ceteris paribus, whether player A delegates to a dice or to player B. The intention (i.e. the unkindness) of delegating is constant across the treatments *D&P* and *random*. In both cases, player A does not choose the fair allocation but delegates the decision, which results in the unfair allocation with a probability of 0.4. The notion of intention based reciprocity then predicts (assuming correct beliefs) the same punishment for player A in both treatments. Regarding player B's punishment in treatment *random* in the situation in which player A delegates, both the notion of intention and responsibility predict no punishment.

The measure of responsibility is based upon a player's share in the probability increase that the unfair allocation realizes. In treatment *asymmetric*, we vary player A's and B's relative impact as compared to treatment *D&P*. If player A cannot be unfair, then by delegating the decision right to player B, he increases the probability that the unfair outcome will result by more

[22] To avoid experimenter demand effects we did not inform the participants that we had previously conducted treatment *D&P* and that the dice matches the probabilities with which player Bs chose between the two allocations.

than in treatment *D&P*.[23] We thus predict that in case the unfair allocation results after delegation, player A will be punished more for delegating in treatment *asymmetric* than in treatment *D&P*. Since shifting responsibility to another player is a zero sum game, the prediction for player B's punishment is exactly opposite. The notion of intentions makes a slightly different prediction in this treatment. In treatment *asymmetric*, delegation is the unkindest action, because the choice of the unfair allocation is not available any longer. If player B is unfair after delegation, then this is also player B's unkindest action. The notion of intention based reciprocity thus predicts that players A and B are punished equally.

Punishment Patterns.—Figure 6 shows the average punishment points that were assigned to players A, B, and the respective other player C in the different situations of our control treatments. The exact values can be read from Table 2 below. The control treatments are designed to study the change in player Cs' punishment assignments in the situation in which player A delegates and the unfair allocation results, relative to the benchmark treatment *D&P*. Notice first that in the other situations, our control treatments replicate the basic findings from treatment *D&P*. In both treatments, *random* and *asymmetric*, if player A makes the allocation choice himself, then there is almost no punishment if the fair allocation is chosen, and if in treatment *random* player A chooses the unfair allocation, then essentially only player A is punished. We also observe, again, that player A's punishment after delegation depends on the subsequent allocation choice. In both cases, after delegation to the dice and to player B, player A is punished significantly less if the fair allocation results than if the unfair allocation results; two-sided Wilcoxon signed rank tests ($p<0.01$).[24]

[23] This holds unless sufficiently more player Bs choose the fair allocation in treatment *asymmetric* than in treatment *D&P*, which was not the case as is reported below.

[24] As argued before, this is not in line with an intention based punishment motive because player Cs' beliefs about player A's intention cannot depend on the final allocation. This is immediately evident in treatment *random*, were the dice's randomization probability and thus Player A's belief about the consequences of delegation are common knowledge at the onset of the game. For treatment *asymmetric* see footnote 12.

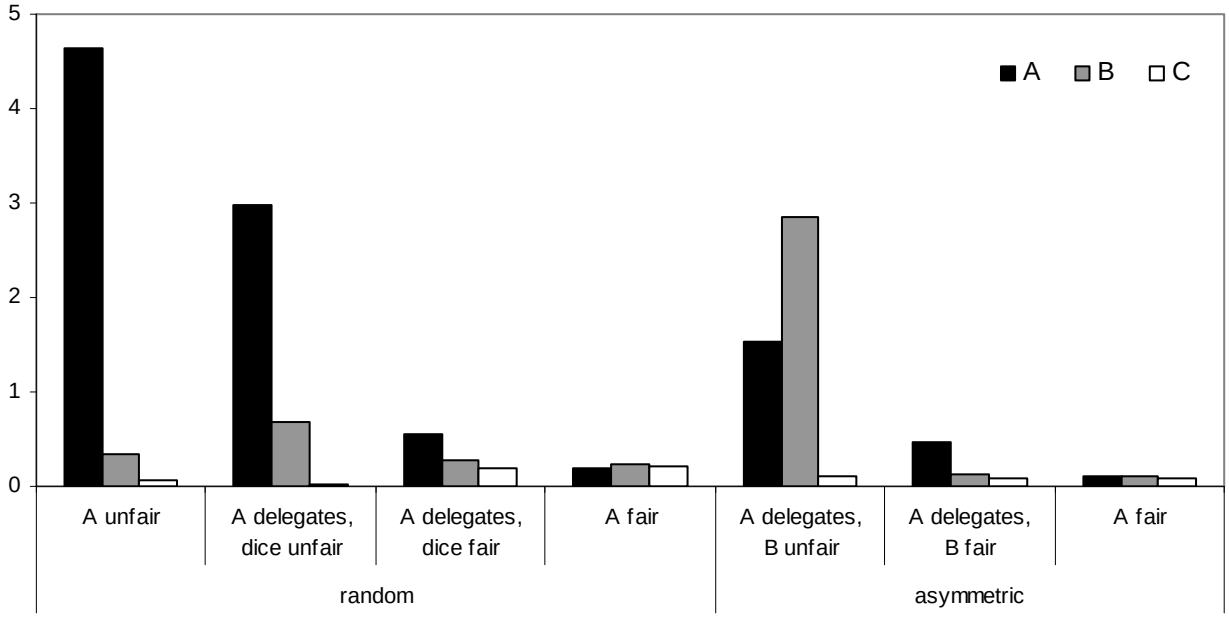

FIGURE 6. PUNISHMENT PATTERN IN THE CONTROL TREATMENTS

The first important finding is that the punishment patterns in the situation in which player A delegates and the unfair outcome results are strikingly different in the treatments *random* and *D&P*. Player A is punished significantly more in treatment *random* if he delegates to the dice and the dice is unfair (2.98 points) than in treatment *D&P* if he delegates to player B and player B is unfair (1.31 points); two-sided rank sum test ($p<0.01$). Moreover, in this situation player A is the player who is punished most; two-sided rank sum tests ($p<0.01$). This observation supports the prediction that responsibility cannot be shifted to a random device; it does not support the prediction of intention based reciprocity, which predicts no treatment difference. However, player A is punished significantly less if he delegates to the dice and the unfair allocation results (2.98 points) than if he chooses the unfair allocation directly (4.64 points); two-sided Wilcoxon signed rank test ($p<0.01$). The punishment patterns thus show that although player A is considered to be

the player who is responsible for the unfair outcome, some responsibility alleviation is nevertheless achieved by delegation to the random device.[25]

The second important difference is that in treatment *asymmetric*, we find that player A is punished more if the unfair allocation results after delegation (1.53 points) than in treatment *D&P* (1.31 points). Inversely, in this situation player B is punished less in treatment *asymmetric* (2.85 points) that in treatment *D&P* (3.96 points). While this is in line with the comparative static prediction of the measure of responsibility, the difference is significant only for player B (rank sum test, $p=0.012$) but not for player A (rank sum test, $p=0.46$). Notice finally that the prediction of the intention measure is not confirmed. Even though both player A and player B took the unkindest action and should thus be punished equally hard, player B is punished significantly more than player A; two-sided Wilcoxon signed rank test ($p<0.01$).

Frequency of Delegation and Allocation Choices.—The choice patterns in the treatments *D&noP* and *D&P* showed that blame shifting is an important motive for the delegation of a decision right. Comparing the punishment patterns in treatments *D&P* and *random*, we find that player As succeed less in avoiding punishment by delegating the decision right to the dice. Consistent with this finding, only 39% of player As (13 out of 33) delegate in treatment *random* compared to 55% in treatment *D&P*. While this difference in the share of delegated decision is in line with our proposition, it is not significant (one-sided Fischer exact test, $p=0.14$). In treatment *asymmetric*, even though player As' average punishment is slightly (but not significantly) higher

[25] In a related paper, Blount (1995) studies the effect of causal attributions on social preferences. She finds that small ultimatum game offers are accepted more often by the receiver if the offer was made by a random device rather than by an intentional agent. In our treatment random, player B is comparable to Blount's "proposer" if the split is determined by the random device. We find that player B is almost not punished if the dice selects the unfair allocation, which is in line with Blount's result. The central question of our paper is, however, different because we are mainly interested in studying whether player A can shift the punishment by intentionally delegating to the dice. In the domain of positive reciprocity, Charness (2000) finds that workers respond with more generosity in a gift-exchange experiment when wages are determined by a random process than when assigned by a third party. He argues that according to the principle of responsibility alleviation, a subject in the random treatment cannot avoid accepting full responsibility for the final allocation, while a high wage that is assigned by a third party may be perceived to be a personally sanctioned entitlement.

as compared to treatment *D&P*, we find that 78% of player As delegate (28 out of 36).[26] This share is marginally significantly different from the share of delegated decisions in treatment *D&P* (Fischer exact test, one-sided p=0.04; two-sided p=0.08). However, the treatment *asymmetric* is by design not directly comparable to the treatment *D&P*, because player As who want the unfair allocation to be chosen must now delegate. Looking at expected payoffs, we find that in both treatments delegation was the most profitable choice. Consistent with this observation, the majority of player As delegated the decision right.[27]

Exact Values of the Responsibility and Intention Measures.—In our belief elicitation session we also asked for beliefs in our control treatments, which allows again to derive exact measures of responsibility and unkindness. The numbers in Table 2 are calculated on the basis of the belief elicitation session.[28]

TABLE 2—ASSIGNED PUNISHMENT POINTS AND PUNISHMENT MOTIVES IN CONTROL TREATMENTS

		random				*asymmetric*		
		A unfair	dice unfair	dice fair	A fair	B unfair	B fair	A fair
average punishment	A	4.64	2.98	0.56	0.18	1.53	0.47	0.11
	B	0.35	0.68	0.27	0.23	2.85	0.13	0.10
responsibility	A	1	0.50	0	0	0.20	0	0
	B	0	0	0	0	0.80	0	0
intention	A	1	0.40	0.40	0	1	1	0
	B	0	0	0	0	1	0	0
outcome	A	1	1	0	0	1	0	0
	B	1	1	0	0	1	0	0

[26] Of the 28 players B who could take a decision in treatment asymmetric, 10 chose the fair allocation.
[27] In treatment *asymmetric*, delegation resulted in an expected payoff of 6.42 while the choice of the fair allocation resulted in an expected payoff of 4.89 (Wilcoxon signed rank test, p<0.01). In treatment *random*, delegation leads to an expected payoff of 5.07, while the choice of the fair and unfair allocation resulted in expected payoffs of 4.82 and 4.36, respectively. These differences are only significant at the 5 and 10 percent level, respectively; two-sided Wilcoxon signed rank test, p=0.02 (fair vs. delegation) and p=0.07 (unfair vs. delegation).
[28] As in treatments *noD&P* and *D&P*, the numbers are derived by averaging over the values of the responsibility measure that result from the participants' beliefs (see also footnote 21).

Table 2 shows the average punishment in the different situations in our control treatments together with the predictions of the punishment motives responsibility, intention, and outcome. The interesting situations are again the delegated decisions. In treatment *random*, if player A delegates to the dice, then his intention measure is 0.4, because the dice selects the unfair allocation with probability 0.4. The value of 0.5 for player A's responsibility in the situation "dice unfair" results from the fact that exactly half of the participants in the belief elicitation session believed that those player As who do not delegate select the unfair allocation with a probability exceeding 0.4. This results in a responsibility measure of 0, because by delegating a player A decreases the probability that the unfair outcome will result. All other beliefs result in a responsibility measure of 1, because a player A is then the only human player that increases the probability that the unfair allocation will result. On average, this yields a responsibility measure of 0.5. In treatment *asymmetric*, if player A delegates to player B and player B is unfair, then both player A's and player B's intention measure equals 1, because both have chosen the unkindest action. The responsibility measure gives much more weight on player B in this situation, which is in line with the punishment pattern. In the next section we use the values in Tables 1 and 2 to test the different punishment motives econometrically.

V. An Econometric Comparison of the Different Punishment Motives

In this section we provide an econometric test of the predictive power of the different motives for punishment. The section is not meant to be a test of different models of social preferences. Rather, we are interested in the predictive power of the responsibility measure, and as a benchmark we use measures that capture the punishment motives in outcome based models and models of intention based reciprocity.

We use regressions to predict the assigned punishment points for the player As and Bs with the three measures that we consider. We also consider the interaction of outcome and intentions. Using the outcome based measure, the regressors equal 1 if the unfair allocation results and 0 if the fair outcome results. Using our responsibility measure, the regressors are calculated as spelled out in equation (3) in Section III. Using the intention measure, the regressors are calculated as spelled out in equation (4) in Section III. The two latter measures depend on beliefs, and we calculate the regressors with the beliefs from our belief elicitation session. They are therefore completely independent of the data of our choice experiments. The exact values can be seen in Tables 1 and 2.

TABLE 3—THE PREDICTIVE POWER OF DIFFERENT PUNISHMENT MOTIVES

	(1)	(2)	(3)	(4)	(5)
outcome	2.041 (0.105)***				0.512 (0.097)***
intention		2.738 (0.161)***			0.314 (0.140)**
outcome*intention			3.189 (0.170)***		-0.194 (0.304)
responsibility				3.76 (0.199)***	3.262 (0.362)***
constant	0.253 (0.043)***	0.286 (0.042)***	0.356 (0.041)***	0.396 (0.041)***	0.213 (0.042)***
observations	1788	1788	1788	1788	1788
R^2	0.21	0.29	0.37	0.42	0.43

Notes: The dependent variables are punishment levels of players A and B. OLS regressions. Robust standard errors are reported in parentheses, allowing for clustering at individual player Cs.
*** denotes significance at 1 percent, ** at 5 percent, and * at 10 percent.

Table 3 shows the results from the regressions. Since all models are translated into one parameter, the R^2 in the regressions (1) – (4) can be compared directly. They provide a clear picture: the predictive power of the outcome motive is lowest (R^2=0.21), the intention motive is second (R^2=0.29), the interaction of outcome and intention is third (R^2=0.37), and the

responsibility motive is clearly best (R^2=0.42). Interestingly, regression (5) shows that outcome and intention add almost no explanatory power on top of responsibility (R^2=0.43).[29]

The results in this section show that the responsibility measure is successful in explaining the punishment pattern (relative to other measures like outcomes or intentions), which lends support to our interpretation that assigned punishment points reflect responsibility attribution.

VI. Conclusion

We conducted controlled laboratory experiments that elicit players' responsibility attributions for delegated decisions. The question of moral responsibility attribution is economically relevant: being responsible means being blameworthy and the prospect of blame can constitute incentives that affect both the actual decision and the decision whether or not to delegate the decision right. If along with the decision right also the responsibility attribution for the ultimate outcome is delegated, then responsibility shirking can be a powerful motive for the delegation of a potential unpopular decision. Our results clearly show that responsibility can be effectively delegated. When an unfair outcome is the result of a delegated decision, then the person who makes the allocation decision is punished much more than the person who delegated the decision. Moreover, the paper demonstrates that responsibility shirking is a powerful motive for the delegation of a decision right. In our main treatment with punishment, 56% of decisions were delegated as compared to 17% in our treatment without punishment.

The results of this paper contribute to a better understanding of why decision rights are delegated. It suggests that the delegation of decision rights is often motivated by reasons other than those usually given in the economics literature (such as skills, work load, commitment, or

[29] Using ordered probit, tobit and fixed effect models, confirms this prediction pattern. Responsibility was always the best single predictor, and in the combined model, outcome and responsibility were always significant, while intention and its interaction with outcome are sometimes insignificant and sometimes even have a negative sign.

incentives). As an example, in the introduction we discussed the business of chief restructuring officers that do not only bring specific knowledge and support to firms undergoing a restructuring process but, as asserted by a company offering such services, also take the blame for unpopular decisions that often have to be made in that process. Our paper shows that such blame shifting indeed works and that decision rights are delegated on exactly this account.

Furthermore, our paper proposes a formal measure of a player's responsibility for an outcome of a game. In a nutshell, the measure attributes most responsibility to players who affected the outcome most. In our experiments, the responsibility measure outperformed outcome and intention based measures in predicting how people assign blame in form of costly sanctions.

Notions of responsibility are important in many contexts beyond the question whether or not to delegate a decision right. Examples include the optimal design of institutions (e.g. hierarchies in organizations) or of incentive systems (e.g. remuneration in team production). How should the scope of actions be designed on different levels in a hierarchy such that ex-post responsibility attribution provides optimal incentives ex-ante? Should explicit monetary incentives mirror ex-post responsibility attribution? Can moral responsibility attributions complement incomplete contracts? Production outsourcing is another example where notions of responsibility are of great importance. The results of this paper suggest that firms may outsource production to shift responsibility, e.g. for poor working conditions. However, usually there is a choice to whom to delegate, which raises many new questions. Does responsibility attribution depend on the selection of the delegee? Does it depend on how much was known about the delegee before the delegation decision was made? Who is blamed if the vulnerability of the parties involved differs? Exploring the implications of notions of responsibility in these contexts constitutes interesting directions for future research.

References

Aghion, Philippe, and Jean Tirole. 1997. "Formal and Real Authority in Organizations." *Journal of Political Economy*, 105(1): 1-29.

Benabou, Roland, and Jean Tirole. 2002. "Self-Confidence and Personal Motivation." *Quarterly Journal of Economics*, 117(3): 871–915.

Blount, Sally. 1995. "When Social Outcomes Aren't Fair: The Effect of Causal Attribution on Preferences." *Organizational Behavior and Human Decision Processes*, 63(2): 131-44.

Bolton, Gary, and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, 90(1): 166-93.

Bolton, Patrick, and Mathias Dewatripont. 2005. *Contract Theory*. Cambridge, MA: MIT Press.

Charness, Gary. 2000. "Responsibility and Effort in an Experimental Labor Market." *Journal of Economic Behavior and Organization*, 42(3): 375-84.

Charness, Gary, and Matthew Jackson. Forthcoming. "The Role of Responsibility in Strategic Risk-Taking." *Journal of Economic Behavior and Organization*.

Charness, Gary, and Matthew Rabin. 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, 117(3): 817-69.

Dana, Jason, Daylian Cain, and Robyn Dawes. 2006. "What you don't know won't hurt me: Costly (but quiet) exit in a dictator game." *Organizational Behavior and Human Decision Processes*, 100(2): 193-201.

Dana, Jason, Roberto Weber, and Jason Kuang. 2007. "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness." *Economic Theory*, 33(1): 67-80.

Dufwenberg, Martin, and Georg Kirchsteiger. 2004. "A Theory of Sequential Reciprocity." *Games and Economic Behavior*, 47(2): 268-98.

Falk, Armin, and Urs Fischbacher. 2006. "A Theory of Reciprocity." *Games and Economic Behavior*, 54(2): 293-315.

Fehr, Ernst, and Klaus Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114(3): 817-68.

Fershtman, Chaim, and Uri Gneezy. 2001. "Strategic Delegation: An Experiment." *RAND Journal of Economics*, 32(2): 352-68.

Fiorina, Morris. 1982. "Legislative Choice of Regulatory Forms: Legal Process or Administrative Process?" *Public Choice*, 39(1): 33-66.

Fiorina, Morris. 1986. "Legislator Uncertainty, Legislator Control and the Delegation of Legislative Power." *Journal of Law, Economics and Organization*, 2(1): 133-51.

Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-made Economic Experiments." *Experimental Economics*, 10(2): 171-78.

Forsythe, Robert, Joel Horowitz, Nathan Savin, and Martin Sefton. 1994. "Fairness in Simple Bargaining Experiments." *Games and Economic Behavior*, 6(3): 347-69.

Greiner, Ben. 2004. "An Online Recruitment System for Economic Experiments." In *Forschung und wissenschaftliches Rechnen*, ed. Kurt Kremer and Volker Macho, 79-93. Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung.

Hamman, John, George Loewenstein, and Roberto Weber. 2007. „Self-interest through agency: An alternative rationale for the principal-agent relationship." Carnegie Mellon University Working Paper.

Herring, Pendleton. 1940. *Presidential Leadership***.** New York: Farrar & Rinehart.

Huck, Steffen, Wieland Müller, and Hans-Theo Normann. 2004. "Strategic Delegation in Experimental Markets." *International Journal of Industrial Organization*, 22(4): 561-74.

Jones, Stephen. 1989. "Have Your Lawyer Call My Lawyer: Bilateral Delegation in Bargaining." *Journal of Economic Behavior and Organization*, 11(2): 159-74.

Konow, James. 2000. "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions." *American Economic Review*, 90(4): 1072-91.

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*, 1(3): 593-622.

Machiavelli, Niccolo. 1961. *The Prince*. London: Penguin Classics.

Manove, Michael. 1997. "Job Responsibility, Pay and Promotion." *The Economic Journal*, 107(440): 85-103.

Prelec, Drazen, and Ronit Bodner. 2003. "Self-signaling and self-control." In *Time and Decision*, ed. George Loewenstein, Daniel Read, and Roy Baumeister, 277-98. New York: Russell Sage Press.

Prendergast, Canice. 1995. "A Theory of Responsibility in Organizations." *Journal of Labor Economics*, 13(3): 387-400.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83(5): 1281-1302.

Rogoff, Kenneth. 1985. "The Optimal Degree of Commitment to an Intermediary Monetary Target." *Quarterly Journal of Economics*, 100(4): 1169-89.

Schelling, Thomas. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Schotter, Andrew, Wei Zheng, and Blaine Snyder. 2000. "Bargaining Through Agents: An Experimental Study of Delegation and Commitment." *Games and Economic Behavior*, 30(2): 248-92.

Sliwka, Dirk. 2006. "On the Notion of Responsibility in Organizations." *Journal of Law, Economics, and Organization*, 22(2): 523-47.

Vaubel, Roland. 1986. "A Public Choice Approach to International Organization." *Public Choice*, 51(1): 39-57.

Vickers, John. 1985. "Delegation and the Theory of the Firm." *The Economic Journal*, 95(Conference Papers): 138-47.