Experimental Economics (2024) 27:637–663
https://doi.org/10.1007/s10683-024-09833-1

# Blame and praise: responsibility attribution patterns in decision chains

**Deepti Bhatia**[1,2] · **Urs Fischbacher**[1,2,3] · **Jan Hausfeld**[4,5] · **Regina Stumpf**[1,2]

Received: 9 May 2022 / Revised: 10 June 2024 / Accepted: 15 June 2024 /
Published online: 9 July 2024
© The Author(s) 2024

## Abstract

How do people attribute responsibility when an outcome is not caused by an individual but results from a decision chain involving several people? We study this question in an experiment, in which five voters sequentially decide on how to distribute money between them and five recipients. The recipients can reward or punish each voter, which we use as measures of responsibility attribution. In the aggregate, we find that responsibility is attributed mostly according to the voters' choices and the pivotality of the decision, but not for being the initial voter. On the individual level, we find substantial heterogeneity with three overall patterns: Little to no responsibility attribution, pivotality-driven, and focus on choices. These patterns are similar when praising voters for good outcomes and blaming voters for bad outcomes.

**Keywords** Responsibility attribution · Collective decision-making · Voting · Decision process

**JEL Classification:** C91 · C92 · D63 · D70 · D91

We thank the audience at the TIBER Symposium 2019, GfeW 2021, M-BEES 2022 and BREW-ESA 2023, the research group of the Thurgau Institute of Economics, and Erte Xiao for fruitful discussions and helpful comments. We gratefully acknowledge support from the German Research Foundation (DFG) through research unit FOR 1882 "Psychoeconomics". The replication material for the study is available under 10.48606/NtWVnKNIdhewVqcw (DOI). The authors have nothing to declare (see Appendix A.13). This research project is in line with the guidelines of the Ethics Committee of the University of Konstanz (see Appendix A.13).

Extended author information available on the last page of the article



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

## 1 Introduction

How is responsibility attributed when an outcome results from a chain of actions? First, consider a disaster (i.e., sinking of large ships) as an example of a bad outcome. Disasters are often a result of a chain of unfortunate circumstances, decisions, and actions. Whittingham (2004) presents several examples of disasters and discusses the responsibility of the different people involved. Typically, someone makes a mistake which is not detected or appropriately fixed. This mistake causes/adds to further problems until a disaster is unavoidable. Similarly, good outcomes are often the result of the (sequential) interaction of people. For instance, joint production is an example of a positive decision chain. If a firm releases a product, research and development, production, and marketing sequentially contribute to the success. In our study, we experimentally investigate how people attribute responsibility in decision chains by allowing people to allocate blame and praise to others in the form of punishment and reward.[1]

The general question of responsibility attribution has been addressed from different angles. A normative point of view has been taken from a philosophical (Feinberg, 1970) as well as from a legal perspective (Hart and Gardner, 2008). More recently, the question has also attracted the interest of psychologists (Ross and Nisbett, 1991; Weiner, 1995; Gerstenberg et al., 2011) political scientists (Iyengar, 1994) and economists, both from an empirical (Charness, 2000; Bartling and Fischbacher, 2012; Bartling et al., 2015; Duch et al., 2015) and theoretical perspective (Besley, 2006; Bartling and Fischbacher, 2012; Engl, 2022). Our study has an empirical point of view as we investigate in a lab experiment how people assign responsibility using punishment and reward. Understanding the empirical patterns of responsibility attribution is important because it has consequences for how we setup liability rules and how we distribute the benefits of joint ventures. Our experiment shows what rules of responsibility attribution people spontaneously apply, and the analysis of the heterogeneity provides insights on how well people agree with respect to these rules.

In real life decision chains, the decision makers and the actions differ in many dimensions. In our experiment, we study the impact of the sequence in isolation. For this reason, we investigate decisions in a sequential voting game, in which symmetric voters decide with majority over a good or bad outcome for other people. The subjects in our experiment are matched into groups of five voters and

[1] Generally, responsibility is associated with blame and praiseworthiness. According to the Oxford dictionary (https://www.oxfordlearnersdictionaries.com/definition/english/responsibility, last retrieved: 06.03.2024), responsibility comes with three meanings: 1. a duty to deal with or take care of somebody/something, so that you may be blamed if something goes wrong. 2. responsibility (for something); blame for something bad that has happened. 3. a moral duty to do something or to help or take care of somebody because of your job, position, etc. The second meaning and the second part of the first meaning relate responsibility attribution to blame attribution. For this reason, punishment and reward is often used in experiments in order to assess responsibility attribution (Coffman, 2011; Bartling and Fischbacher, 2012; Duch et al., 2015; Gurdal et al., 2013; Oexl and Grossman, 2013; Bartling et al., 2015). There are also other ways to assess responsibility attribution, for example, Engl (2022) directly asks participants about the responsibility of different actors in different scenarios using the Krupka and Weber (2013) method.



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

five recipients. The five voters choose sequentially between two options of how to allocate points between voters and recipients, while the outcome is determined by simple majority rule. The two allocations differ in their fairness: the unfair allocation favors the voters, while the fair allocation results in similar payoffs for the voters and recipients. The recipients receive full information about the voting sequence and then have the possibility to sanction the voters. Our treatments differ in the sanctioning options. There is a treatment with only punishment, one with only reward, and one with both. Finally, we use process measures and record the response times for all participants and use eye-tracking for the recipients. We make several contributions to the existing literature. We study different outcomes (good and bad outcomes), compare different sanctioning options (punishment and reward), explore more motives (outcome, choice, intention, initiation, pivotality, causal responsibility based on models), and add process measures (response time and eye-tracking). We are investigating the following research questions:

First, we study *how responsibility is attributed to different roles in the decision chain*. We investigate how the position in the decision chain affects responsibility attribution, and we use two theoretical measures of responsibility as predictors (Bartling and Fischbacher, 2012; Engl, 2022). Two voters are in a particular focus, the first voter in favor of the resulting outcome, which we call initiator and the third voter in favor of the resulting outcome, which we call the pivotal voter.[2]Bartling et al. (2015) show that in sequential decisions pivotal voters are blamed the most for unfair outcomes. Duch et al. (2015) find that in simultaneous collective decisions, proposal power plays an important role for responsibility attribution. In our design with five voters, we can compare the initiator and the pivotal player with a majority voter who is neither. Further, we can distinguish between majority voters who still had a say and those who had no more influence on the outcome. We call the former the intentional voters and the latter the non-intentional voters.[3] On the aggregate, our experiment shows that punishment targets people who vote for the unfair outcome. People who potentially had an impact on the outcome, and in this respect are intentionally unkind, are punished more, and the pivotal players even more within this group. Analogously, choosing the kind option leads to rewards, which are higher if the choice can be considered as intentional and even higher if the choice is pivotal. We do not find that higher reward or punishment is assigned to the initiator of a good or bad outcome. The theoretical measures of responsibility (Bartling and Fischbacher, 2012; Engl, 2022) are correlated with the empirical responsibility attribution but the correlations are not very high and these measures do not outperform simpler measures.

[2] The term pivotality is used in a variety of contexts. In voting decisions, pivotality is also used with respect to the order that is induced by the strength of the preferences in favor of a policy. We use the term with respect to the temporal order. This is meaningful because after the third decision in favor of an option, the outcome of the vote is determined. Our definition also corresponds to the definition in Bartling et al. (2015). A definition of gradual pivotality is used by Engl (2022) as a measure of causal responsibility. We discuss this definition in Sect. 3.

[3] If the decision is already taken, we cannot infer any intention of the voter. For example, if the outcome has already been decided, a vote for the fair outcome does not mean that the player wanted to be kind.



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

Second, we study *whether responsibility attribution differs for good and bad outcomes in comparable situations.* We do so, by making use of our treatments with reward and punishment options, as well as the combination, in which both reward and punishment are available. As mentioned above, the evaluation of responsibility is consistent between reward and punishment. Subjects reward others for good outcomes very similarly to how they punish others for bad outcomes – both on the aggregate level as well as in the individual analysis introduced below. The environment in which both reward and punishment are available shows that people tend to prefer to use punishment over reward. However, responsibility attribution is less differentiated compared to the environments in which only one option is available.

Third, as the explanatory power of the general model is not very high, we explore *individual patterns in responsibility attribution*. Studying different patterns in social interactions has been the object of several studies.[4] We find that the individual behavior can be classified into three main groups. These groups are analogous in the reward and the punishment condition. There is a group of subjects who barely rewards or punishes. Another group of subjects particularly targets the pivotal voter, and a third group of subjects mostly attributes responsibility according to the choices of the voters.

Fourth, we investigate *how the voters respond* to the incentives created by the option of punishment and reward. We find evidence that voters are (at least partially) aware of how responsibility is attributed. In particular, they are aware that pivotality matters and partially use delegation in order to avoid blame for unkind decisions or seek responsibility for kind decisions in order to gain credit.

Fifth, we study *the underlying decision process of responsibility attribution.* We analyze the response time patterns of voters and find that they have longer response times when they are potentially pivotal, i.e., if their decision can finalize the outcome. Further, we use eye tracking during the sanctioning decisions. Our results show that the gaze analysis does not confirm the behavioral focus on the pivotal player. If any player is more in the focus, then it is the initiator.

The remainder of this paper is structured as follows: Sect. 2 explains the experimental design used in this study while Sect. 3 outlines the different motives used to study responsibility attribution and lists our predictions. Our results are presented in Sect. 4 and Sect. 5 concludes.

[4] For example, Falk et al. (2008) investigate different patterns of reward and punishment among their subjects. Most participants express both positive and negative reciprocity, while others only show positive or negative reciprocal fairness preferences. Similarly, Leibbrandt and López Pérez (2011) study heterogeneity in costly reward and punishment. Their results indicate that most subjects follow a mixture of outcome-based and reciprocal preferences. Besides observing different patterns of how subjects sanction, Albrecht et al. (2018) go one step further and examine if and how different behavioral patterns are linked within each subject. In a linear public goods game with decentralized punishment, they show that for most subjects cooperation and punishment patterns are aligned.



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

## 2 Experimental design

We build on Bartling et al. (2015) who investigated a sequential voting task with punishment. We add a treatment with reward in order to directly compare responsibility attribution for good and bad outcomes, and we increase the number of voters as it allows investigating more roles in this sequential decision process. In our experiment, we randomly assign the role of voter and recipient to subjects. Five voters and five recipients form a group and keep their roles throughout the experiment. The five voters sequentially decide between two allocations in order to distribute 50 points among all ten group members. There are two sets of allocations. In one set, voters can choose between a fair allocation, in which all group members receive 5 points, and an unfair allocation, in which the voters receive 9 points each and the recipients receive 1 point each. In the second set, the voters can choose between a fairer allocation, in which the voters receive 6 points and the recipients receive 4 points, and an unfair allocation, in which the voters receive 8 points each and the recipients receive 2 points each. We chose the two sets in such a way that the alternatives create a similar trade-off, and, thus, the two sets can be treated equally.

The position of the voters in the voting sequence is randomly determined. Each voter is informed about the decisions of all previous voters in the sequential process before choosing an allocation. A majority rule is applied, which means that the allocation that is chosen by at least three of the five voters is implemented.

The recipients are informed about the individual voting decisions and thus also about the voting outcome. One randomly determined recipient receives an extra point and has the option to sanction the voters. We vary the sanction option across three treatments: Recipients can only punish voters (*Punishment*), they can only reward voters (*Reward*), and they can reward and punish voters (*Both*). In all the treatments, the recipient first has to decide whether to sanction the voters at the cost of the extra point by clicking a button. In the second decision, the recipient can then assign 0 to 7 reward and/or punishment points to each voter individually. Figure 1 illustrates an exemplary decision screen of a selected recipient who decided to sanction the voters.

The payoff of each voter is determined by the resulting voting outcome and the reward or punishment points the voter receives from the recipient. Each recipient gets a payoff according to the chosen allocation. The selected recipient can additionally keep the extra point if she decides not to sanction.

The game is played as a one-shot game and we use the strategy method for both voters and recipients, that is, each voter and each recipient makes choices for all possible scenarios. Each voter chooses between the fair and unfair allocation in every voter position for every possible combination of previous voter choices. This results in 31 binary choices for each of the two allocation sets which we display in random order. Additionally, the voters play one round of a dictator game for each allocation set resulting in two additional decisions. All recipients act as if they were chosen to be the recipient to sanction. For every possible voting sequence, the recipients decide whether they want to sanction any of the



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

voters and if yes, by how much they want to sanction each voter. The scenarios differ in the decision constellations of the voters and the allocation sets.

As process measures, we collect response time data for both voters and recipients. In addition, we use eye-tracking to record the gaze pattern of the recipients to evaluate the information recipients use when attributing responsibility to the voters.

### 2.1 Procedural details

The experiment was programmed using the software "z-Tree" (Fischbacher, 2007). Participants were students who were recruited by the data-system ORSEE (Greiner, 2015). In total, nine sessions were conducted in February 2019, three sessions for each treatment. The experiment was carried out at the experimental laboratory of the University of Konstanz (Lakelab) in Germany. Each session consisted of two groups, 10 voters and 10 recipients, such that there were 30 voters and 30 recipients in each of the three treatments (*Punishment*, *Reward*, *Both*). One subject was excluded from the analysis due to insufficient attention during the session.

The average age of our subjects was 22 years (min: 18, max: 33) and 55.56% of the subjects were female. The subjects earned on average 22.58 EUR (about 25.40 USD at that time) which included a show-up fee and an extra compensation for the usage of eye-tracking. The sessions lasted 90 min on average.

We used Tobii EyeX eye-trackers with a sampling rate of 60Hz to record gaze data. The subjects used chin rests to improve data quality and the seating distance to the screen was approximately 58 cm. The screens were 22 inch color monitors with a resolution of 1920x1080 pixels. The calibration of the subjects to the eye-tracking system was done at the beginning of the experiment via a seven-point calibration. Two additional subjects have been excluded from the eye-tracking analysis because of technical issues and poor quality of the gaze data.

Fixations are identified with the help of the DBSCAN-algorithm (Ester et al., 1996). We create ten non-overlapping areas of interest (AOI), each with a radius of 90 pixels. Each AOI covers a box on the decision screen indicating whether the voter voted in favor or against the specific allocation (see Fig. 1). Therefore, for each voter there are two AOIs. The horizontal distance between the centers of two AOIs was 320 pixels and the font size of the cues was set to 20.

## 3 Criteria and theoretical predictions

In this section, we present criteria according to which people could attribute responsibility. It includes motives like intention-based reciprocity and measures of responsibility. We will assess the relevance of these criteria by using them as predictors for reward and punishment. The measures that we will present first have not been intended to be used as measures of responsibility. However, these measures have been suggested to explain reward and punishment. Even when they do not capture all facets of responsibility, they capture some facets, which has also been



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

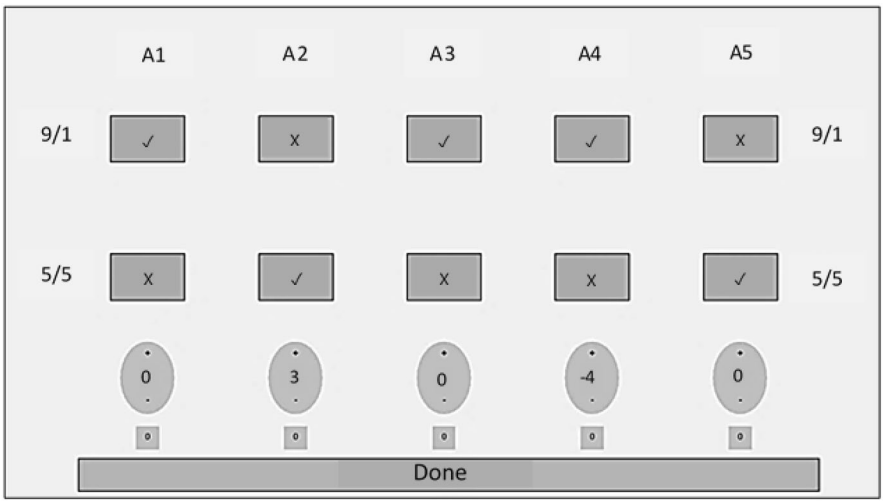



**Fig. 1** Exemplary decision screen of a recipient in the treatment *Both*. Note: Voters are denoted as A1 - A5 and their decisions are indicated by a check at the selected allocation. The positioning of the allocations in the top or bottom row was randomly determined. The outcome in this example is the unfair allocation (9 points for each voter and 1 point for each recipient) and the recipient attributes three reward points to voter 2 and deducts four punishment points from voter 4. We added the respective allocation on both sides of the screen to minimize subject's gaze being biased towards one side of the screen. The font size in the figure was enlarged for better readability

discussed in the theoretical literature on responsibility. For example, Shaver (1985) mentions dimensions of responsibility, among them, intentions.

**Outcome.** Outcome-oriented models such as Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) are quite popular. In our voting game, it means that reward and punishment is not directed to a specific voter. The assessment only depends on the outcome. The unfair alternative is considered as bad, and the fair alternative as kind or neutral.[5] The models predict equal punishment for all voters in case of an unfair outcome and if at all, equal reward in case of a fair outcome.

**Choice.** This motive assumes that a vote for the fair allocation is perceived as kind and a vote for the unfair allocation as unkind, independent of whether the vote was relevant for the outcome or not. It can be considered as a naive notion of intention. If the voter would not take the behavior of the other voters into account and would believe that their own vote is decisive, then their own vote would correspond to their preference and the vote would express their intention.

**Intention.** This motive captures preferences as suggested in the reciprocity models of Rabin (1993) and Dufwenberg and Kirchsteiger (2004). Since there are only two options, either both are neutral, or one option is kind and the other is unkind. Theoretically, whether voting for fair or unfair is kinder depends on the belief of

[5] According to outcome-oriented models, voting for the fair alternative can be considered as neutral and therefore fair outcomes might not be rewarded.



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

what the other voters do.[6] However, in the experiment, voting for the fair allocation always results in a higher probability of getting the fair outcome than voting for the unfair allocation. Thus, as long as a majority for one of the allocations is not reached, votes are impactful and voting for the fair allocation can be considered as kind and voting for the unfair allocation as unkind. As soon as the decision is made and the outcome can no longer be changed, votes are no longer impactful, no intentions can be inferred, and the vote is considered as neutral.

**Initiation.** This motive is motivated from Duch et al. (2015) who showed that proposal power is an important aspect in how subjects attribute responsibility for collective decisions.[7] Applied to our experiment, it assumes that the first voter who votes for the resulting outcome has a special responsibility for the outcome.

**Pivotality.** This motive is motivated from Bartling et al. (2015) who found that the pivotal voter is punished more than the non-pivotal voters for voting for the unfair outcome. In our setup, the pivotal voter is the third voter who votes in favor of one of the two allocations. After this choice, the outcome is determined and can no longer be changed.

**Bartling and Fischbacher** (2012) **Responsibility Measure.** In the responsibility measure formalized by Bartling and Fischbacher (2012) (from now on called BF Responsibility), the responsibility of the different voters for a certain outcome is assigned proportional to how much their vote contributes to an increase in the probability that this outcome results. The measure depends on the belief about the voter's decisions. In order to keep the following explanations simpler, we present the case where the outcome is unfair. The measure works exactly in the same way for the fair outcome. It is calculated as follows: First, we calculate for every decision node the probability that the unfair outcome results. Next, each action (e.g. each vote) gets a raw responsibility, which is the difference between the probability before and after the action. Finally, in order to get the responsibility measure, the raw responsibility is normalized. This means specifically: The responsibility of an action that does not increase the probability of an unfair outcome is set to zero. The responsibility of an action that increases the probability of an unfair outcome is the raw responsibility divided by the sum of all positive raw responsibilities along the decision path from the start to the final outcome. This measure lies between zero and one. As mentioned above, it depends on a belief about the voters' decisions. Practically, we use the empirical distribution of voters' decisions as their belief. Note that the responsibility measure refers to the outcome. We expect that responsibility for the unfair outcome triggers punishment and responsibility for the fair outcome triggers reward.

**Engl** (2022) **Responsibility Measures.** Another notion of responsibility has been suggested by Engl (2022). We consider a simple variant of the model and explain what it predicts in our case. Engl (2022) distinguishes between ex-ante and ex-post causal responsibility. For the calculation of ex-post causal responsibility,

[6] For example, if voters 2 to 5 always vote against voter 1, then the fair vote of voter 1 would actually be unkind.

[7] However, proposal power is not explicit in our experiment. This may limit how people assess the initiator's responsibility.



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

the outcome is considered as given. The idea of this measure is that the attributed responsibility of an action increases in proportion to the pivotality of this action. An action is considered as pivotal if it causes the outcome and not choosing it would result in the counter-factual outcome. With respect to the ex-post responsibility, the action has the outcome as consequence and, therefore, the responsibility of the action is just the probability that not choosing this action causes the other outcome.[8] Note that the ex-post responsibility is defined for an action within a decision path. The ex-ante responsibility of a vote is calculated as the expected value of the ex-post responsibility of all paths following this vote. Thus, the ex-ante responsibility fixes the target vote as well as the preceding votes and calculates the expected value of the ex-post responsibilities over all paths that follow the target vote.

Let us illustrate the idea of ex-post responsibility in a few examples where F refers to a vote for the fair allocation, U refers to a vote for the unfair allocation, and the order of F and U refers to the sequence of votes. So, FUFUU represents a situation, in which the first and the third voters vote for fair and the other voters vote for unfair. The fourth and the fifth vote in this sequence are fully responsible for the unfair outcome because choosing fair would result in the fair outcome. In the sequence UFFFF, the fair outcome results. The last voter in this sequence is not responsible for the fair outcome at all because a change of the action does not change the outcome. The ex-post responsibility of the fourth voter in the sequence UFFFF depends on the probability $p$ that the last voter chooses U. If the fourth voter would vote U, then the unfair outcome would result with a probability of $p$. Thus, the ex-post responsibility of voter four for the fair outcome equals $p$. Note that different from the BF responsibility measure, the Engl responsibility for the fair outcome is not always zero when the unfair outcome results and vice versa. For example, the first unfair vote (U) in UFFFF does not have zero responsibility for the fair result, and the first voter who votes for F bears some positive responsibility for the unfair outcome. Therefore, we consider the difference between the responsibility for the unfair and the fair outcome as predictors. We use the variables *Ex-ante Engl Difference (U-F)* and *Ex-post Engl Difference (U-F)* as predictors for the *Punishment* treatment and *Ex-ante Engl Difference (F-U)* and *Ex-post Engl Difference (F-U)* as predictors for the *Reward* treatment.

Table 1 shows a summary of the presented criteria and their theoretical predictions, which we use to analyze how people attribute responsibility. The third column of the table illustrates the theoretical predictions of how responsibility is assigned for the unfair outcome in the voting sequence FUFUU. The fourth column of the table provides the theoretical predictions of the responsibility assignment for the fair outcome in the voting sequence UFFFF. The tables in Section A.3 of the appendix display the responsibility predictions for each voter in each voting sequence according to the models by Bartling and Fischbacher (2012) and Engl (2022).

[8] This probability also depends on what the other voters subsequently do. We use the empirical distribution of the voter behavior for this purpose.



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

**Table 1** Overview of responsibility measures

| Criterion | Theoretical prediction-Who is responsible? | Responsibility for unfair outcome in | | | | | Responsibility for fair outcome in | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | F | U | F | U | U | U | F | F | F | F |
| Outcome | All voters are responsible for the final outcome | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Choice | All voters who vote for the final outcome are responsible | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| Intention | Impactful voters are more responsible than non-impactful voters | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| Initiation | The first voter voting for the final outcome is responsible | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Pivotality | The third voter voting for the final outcome is responsible | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| BF Responsibility | Voters are more responsible if they increase the probability that the final outcome results | 0 | .29 | 0 | .34 | .37 | 0 | .26 | .51 | .23 | 0 |
| Ex-post Engl Responsibility | Voters are more responsible if choosing the alternative action would have resulted in a different outcome with a higher probability | .18 | .91 | .17 | 1 | 1 | .35 | .93 | .81 | .40 | 0 |
| Ex-ante Engl Responsibility | Expectation of ex-post Engl Responsibility. Fixes the votes preceding the target vote and takes the expectation with respect to subsequent votes | .05 | .57 | .05 | .63 | 1 | .06 | .37 | .66 | .40 | 0 |

Summary of the presented criteria and their theoretical predictions. The third and fourth columns shows the predictions of the criteria for each voter in two voting sequences of fair (F) and unfair (U) choices. The third column reports the responsibility for the unfair outcome in FUFUU. The fourth column reports the responsibility for the fair outcome in UFFFF. BF and the Engl measures depend on the distribution of the voting decisions. We show the measures from the punishment treatment in the allocation set (9, 1) vs. (5, 5)



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

In the exemplary situation of the voting sequence FUFUU, the unfair outcome results. Therefore, the outcome-based models would predict equal punishment for all voters. If Choice is used as a criterion to assign responsibility, unfair choices will be punished equally (second, fourth and fifth voter) and fair choices will be rewarded equally (first and third voter). The intention-based models would predict equal responsibility for the second, fourth and fifth voter for the unfair outcome since a majority of votes was not reached before. With respect to initiation, the second voter would be solely responsible for the unfair outcome since this voter was the first to vote for the unfair allocation. The fifth voter in the sequence is the pivotal voter and would be fully responsible according to this criterion. BF Responsibility would predict zero responsibility for the voters who voted for the fair allocation (first and third position voters) and a responsibility between 0 and 1 for the second, fourth and fifth position voter. As outlined above, the ex-post Engl Responsibility model predicts that the fourth and fifth voters are fully responsible. All the other voters are partially responsible. The ex-ante Engl Responsibility is non-zero for all voters because in all decisions, the unfair outcome is possible. It equals 1 for the last voter, because for the last voter, the ex-ante and the ex-post measures coincide. For voter 4, the ex-ante measure is the expected ex-post measure of the two sequences FUFUF and FUFUU. For voter 3, it is the expected ex-post measure of four sequences, and so on.

## 4 Results

Our main research focus is how recipients use punishment and reward (attribute responsibility) for sequential collective decisions, which are either fair or unfair. To do so, we first test the criteria and theoretical predictions stated in Sect. 3, followed by an integrative model, and a short analysis of individual heterogeneity and process data. In the analysis of the voters' behavior, we study voting patterns, strategic voting and delegation, and response times. We do not distinguish between the two possible allocation settings because the results are, as expected, quite similar.[9] [10]

### 4.1 Sanctioning Behavior

For the analysis, we separate the decisions according to whether the outcome was fair or unfair, and classify each voter into the majority group (those who voted for the resulting outcome) or into the minority group (those who voted against the resulting outcome). In each voting sequence there are between three and five majority voters and between zero and two minority voters. Among the majority voters, we distinguish between voters who vote before a majority is reached and who

[9] In the dictator games, 21 out of 90 subjects chose (5,5) when (9,1) was the alternative and 20 chose (6,4) when (8,2) was the alternative.

[10] The predictions of BF and Engl depend on the voter behavior. Thus, these predictions depend on the treatments. The analysis uses these treatment specific predictions.



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

are therefore impactful for the final outcome (first three majority voters) and non-impactful voters (the possible fourth and fifth majority voter). Finally, we separate the impactful voters (first three majority voters) into the initiator, i.e., first majority voter, the second majority voter and the pivotal voter, who is the third majority voter. The impactful voters are named according to their roles in the sequential decision.

**Outcome.** Fig. 2 shows the average punishment and reward points for fair and unfair outcomes across treatments. Note that in the *Both* treatment, recipients could use the seven points for both punishment and reward. Thus, the *Both* treatment is included in both sub-figures.

In the *Punishment* treatment, recipients punish more in unfair outcomes than in fair outcomes (4.04 and 1.40 points, Wilcoxon signed-rank test, $p < 0.001$).[11] In the *Reward* treatment, more reward points are assigned on average in fair outcomes compared to unfair outcomes (3.63 and 1.59 points, $p < 0.001$). In the *Both* treatment, where subjects can both punish and reward voters, more punishment points are used on average for unfair outcomes (2.84) than for fair outcomes (1.07, $p < 0.001$). In contrast, recipients do not reward fair and unfair outcomes differently (0.91 and 0.83 points, $p = 0.975$). This shows that punishment is used more frequently and in a more differentiated way than reward when both options are available.

Our results show that recipients indeed punish unfair outcomes more than fair outcomes. However, our outcome-based prediction cannot be supported insofar as recipients also punish when the outcome is a fair allocation and use reward points for both outcomes.

**Choices.** Fig. 3 shows the average sanction points for different voter roles for unfair and fair outcomes across treatments. Table A2 in the appendix lists the corresponding average sanction points in more detail by also taking the voter position into account. It is important to note that not every voter role is equally represented across all possible scenarios due to the natural composition of all possible voting choice constellations. Table A3 shows the average sanction points for each voter position in each scenario faced by the recipients.

Figure 3 shows that choice clearly matters for responsibility attribution. In all treatments, voting for the fair allocation is rewarded more and/or punished less than voting for the unfair allocation (all *p-values* $< 0.02$). This is illustrated by comparing the minority voters (yellow bars) with the respective majority voters (blue bars) in Fig. 3 for each outcome. Our choice-based prediction can thus be confirmed by the data and shows that subjects attribute responsibility according to choices. In particular, there is virtually no punishment for fair votes and, vice versa, hardly any reward for unfair votes. Nevertheless, reward points for subjects voting for the unfair allocation as well as punishment points for subjects voting for the fair allocation are significantly different from zero for almost all voter roles (all *p-values* $< 0.05$, except for one case).

**Intentions.** We now disentangle who is held more responsible among the voters choosing the same allocation. We call the first three majority voters of a voting

[11] For every hypothesis test we use a Wilcoxon signed-rank test for matched samples which is based on average decisions per subject.



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

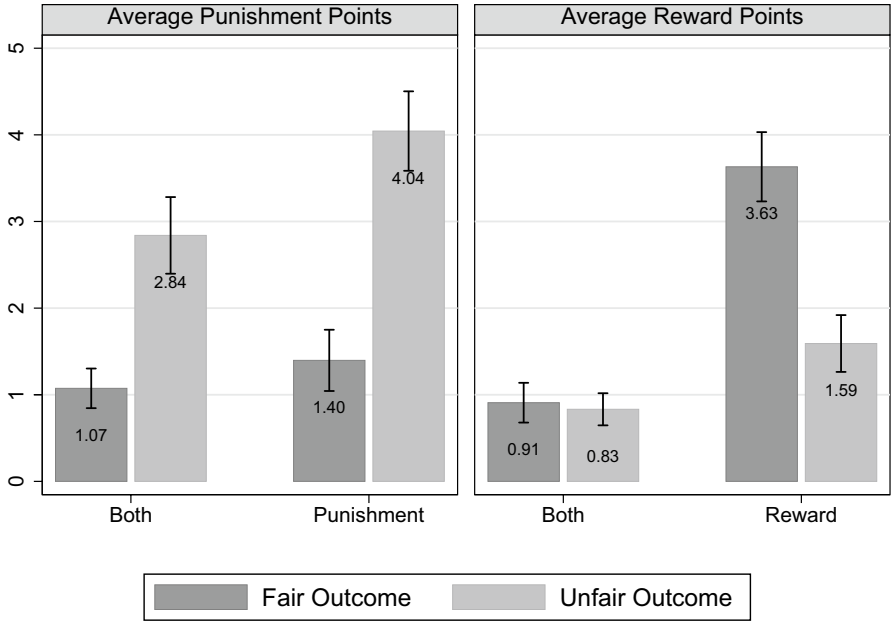



**Fig. 2** Average punishment and reward points for fair and unfair outcomes across treatments. Note: Standard error bars are shown in black

sequence intentional voters, while the majority voters four and five are non-impactful voters as their vote can no longer change the outcome. As shown in Fig. 3, recipients punish intentional voters for unfair outcomes and reward them for fair outcomes more than non-impactful voters (all *p-values* $< 0.05$). In the *Both* treatment, the same results hold except for one case.[12]

Our results show that the intention-based prediction can be confirmed. Thus, when attributing responsibility for an outcome, recipients take the impact of the votes on the final outcome into account.

**Initiation.** We analyze whether there are differences in responsibility attribution among intentional voters. We first test whether the initiator is sanctioned more than the second majority voter. Across all treatments, we do not find evidence for an initiator effect (average sanction points for initiator vs. second majority voter: *Punishment* −1.18 vs. −0.99; *Reward* 0.94 vs. 0.97; *Both* fair outcomes 0.23 vs. 0.22; *Both* unfair outcomes −1.02 vs. −0.72). Recipients do not seem to punish and/or reward the initiator differently than the second majority voter (all *p -values* $> 0.1$). Therefore, the initiation-based prediction does not hold.

**Pivotality.** We expect recipients to attribute the highest responsibility to the pivotal voter (e.g., Bartling et al. (2015)). In the *Punishment* treatment the pivotal

[12] For fair outcomes the second majority voter and the fifth majority voter are not treated differently.



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

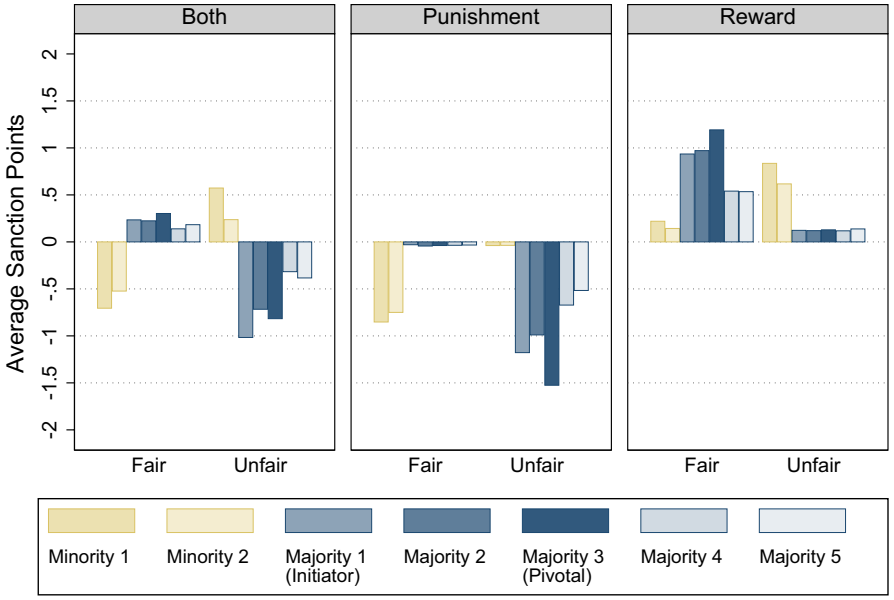



**Fig. 3** Average sanction points for different voter roles across treatments. Note: The bars show the average sanction points for different sanction motives separated by outcome (fair vs. unfair) and treatment. Depending on the voting sequence the following voter roles are possible: Minority 1 represents the first voter voting against the final outcome. Minority 2 represents the second minority voter. Majority 1 (Initiator) is the first voter to vote for the final outcome. Majority 2 is the second majority voter. Majority 3 (Pivotal) is the third voter to vote for the final outcome. Majority 4 and Majority 5 are the fourth and fifth majority voter.

voter is punished more than both other intentional voters when the outcome is unfair ($-1.52$ vs. $-1.18$ / $-0.99$, both *p-values* $< 0.02$). In the *Reward* treatment the pivotal voter is rewarded the most for fair outcomes (1.19 points on average) which is more than the other two intentional voters (both *p-values* $< 0.08$). However, in the *Both* treatment the pivotal voter is not treated differently compared to other intentional voters (all *p-values* $> 0.1$). The pivotality-based prediction can partially be confirmed. Pivotality plays an important role when attributing responsibility for cases where reward and punishment are separately available.

Taken together, our analysis allows us to answer the first two research questions on how responsibility is attributed to different roles in a decision chain and between different outcomes. Our results show that people attribute responsibility differently depending on the outcome of the sequential decision. Generally, subjects are held responsible according to their choices. Furthermore, impactful voters are perceived to be more responsible than non-impactful voters. Last, the pivotal voter bears the highest responsibility while the initiator of a sequential voting sequence is not treated differently than the second impactful voter. In addition, we find that the different criteria of responsibility attribution are the same whether people praise others for good outcomes or blame them for bad outcomes.



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

### 4.1.1 Econometric comparison of sanctioning motives

In this section, we provide an econometric comparison between the different motives and measures of responsibility attribution. As the correlation tables in Section A.4 in the appendix show, there are high correlations between the different motives and measures. Therefore, in order to compare the explanatory power of the different motives, we first study them in isolation.

Figure 4 shows the $R^2$ of individual OLS regressions with reward and/or punishment points as dependent variables and the criteria of responsibility attribution as presented in Sect. 3 as independent variables.[13] The figure shows that the criterion Choice has the highest explanatory power across all treatments, followed by the criterion Intention, further followed by the theories of responsibility attribution (BF, Ex-ante Engl Diff, Ex-post Engl Diff), which have similar explanatory power but among them the Ex-ante Engl Diff measure has the highest explanatory power in all situations. The criteria Pivotality and Initiation are amongst the criteria with the least explanatory power.[14]

So far, each motive of how responsibility is attributed has been tested separately. Outcome-based models predict when people use punishment or reward, but they do not predict who is held responsible. Models based on reciprocity and intentions can explain who is perceived responsible but not when. We now test which motives have explanatory power when considering all motives simultaneously and, thereby, compare the importance of the different motives. Importantly, even though the outcome is the same in many scenarios in our experiment, the number of votes for and against the outcome differs and also which voter position was associated with which motive (i.e., whether the third, fourth, or fifth voter is pivotal; whether the first, second or third voter is the initiator). Table 2 shows the corresponding OLS regression outputs. On the left side of the table we regress the punishment points on each sanctioning motive simultaneously for the treatments *Punishment* and *Both*, while the right side shows the respective regressions for the reward points in the treatments *Reward* and *Both*.[15] Since the measures are correlated, we also report regressions with only the positional variables (choice, outcome, intention, initiator and pivotal), with only the responsibility measures and with single regressors in Appendix A.5. We control for the size of the majority voters but do not include the voter position.[16]

[13] The output of the individual OLS regressions can be found in the Appendix A.5. Note, that when taken individually, each responsibility measure significantly predicts the attributed sanction points.

[14] With respect to explanatory power, the theories of responsibility attribution might be seen as a good compromise between the individual motives. Overall, the explanatory power is not very high and can only explain up to around 18% of the variance. One potential explanation could be individual heterogeneity. We discuss this aspect in Sect. 4.1.2.

[15] We report here the differences between the responsibility for unfair and fair outcomes for the Engl responsibility measures. The other variants can be found in Section A.6 of the appendix.

[16] We tested whether the voter position influences the perceived responsibility of each voter category by regressing the sanction points for each voter role on the voter position. The results show that only for the initiator and the pivotal voter the positioning has an impact on the sanction points. Initiators on position 3 are punished more for bad outcomes than initiators on position 1 and 2. Initiators on position 2 and 3 are rewarded more for good outcomes than initiators on position 1. Pivotal voters on position 5 are pun-



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

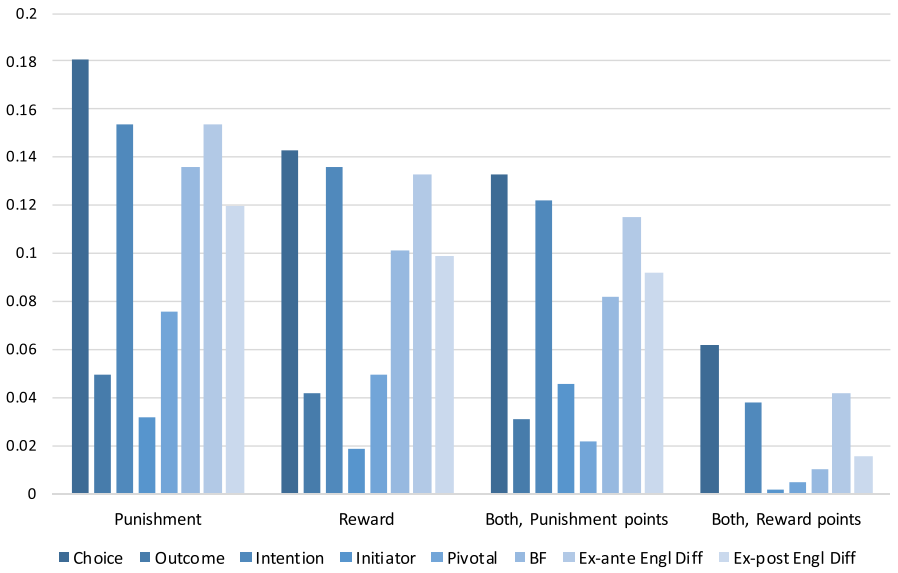



**Fig. 4** Comparison of $R^2$ for different responsibility measures

The regression output in Table 2 shows that choices have an explanatory power on top of all other motives. Unfair choices predict the punishment patterns seen in the treatments *Punishment* and *Both*. In the *Punishment* treatment, voters who choose the unfair allocation get 0.778 more punishment points than voters choosing the fair allocation, and this increase equals 0.466 in the *Both* treatment. On the other side, fair choices are a good predictor for how people reward collective decisions in the treatments *Reward* and *Both*. In these treatments, voters are rewarded 0.455 / 0.335 more reward points when choosing the fair allocation compared to the unfair allocation. In the treatments *Punishment* and *Reward* pivotality has predictive power for the perceived responsibility when considering all motives. Being pivotal for the unfair outcome leads to 0.565 more punishment points and being pivotal for the fair outcome leads to 0.213 more reward points compared to other intentional voters. Looking at the responsibility measures by Bartling and Fischbacher (2012) and Engl (2022) one can see that the criterion Ex-post Engl Difference for fair and unfair outcomes helps in explaining the punishment behavior in this joint regression. The remaining responsibility measures BF Responsibility and Ex-ante Engl Difference do not help much in explaining who is held responsible when combining all measures. This is because the responsibility measures encompass various individual

Footnote 16 (continued)
ished more than pivotal voters on position 4 in unfair outcomes. Pivotal voters on position 5 are rewarded more than pivotal voters on position 3 and 4 in fair outcomes.



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

**Table 2** Joint OLS regressions to compare the impact of the criteria on the usage of punishment and reward points

| | Punishment Points | | | Reward Points | |
|---|---|---|---|---|---|
| | Punishment | Both | | Reward | Both |
| Choice Unfair | 0.778*** | 0.466*** | Choice Fair | 0.455*** | 0.335*** |
| | (0.158) | (0.103) | | (0.111) | (0.084) |
| Outcome Unfair | −0.112 | −0.127* | Outcome Fair | −0.038 | −0.168** |
| | (0.111) | (0.052) | | (0.071) | (0.048) |
| Intention Unkind | −0.022 | 0.070 | Intention Kind | 0.180* | −0.001 |
| | (0.139) | (0.083) | | (0.082) | (0.042) |
| Initiator Unfair | 0.184 | 0.307 | Initiator Fair | 0.005 | −0.043 |
| | (0.228) | (0.242) | | (0.064) | (0.027) |
| Pivotal Unfair | 0.565** | 0.099 | Pivotal Fair | 0.213** | 0.052 |
| | (0.204) | (0.090) | | (0.076) | (0.046) |
| BF Responsibility (U) | 0.478 | 0.187 | BF Responsibility (F) | 0.285 | −0.233 |
| | (0.442) | (0.301) | | (0.348) | (0.167) |
| Ex-ante Engl Diff (U-F) | −0.104 | 0.004 | Ex-ante Engl Diff (F-U) | −0.048 | −0.006 |
| | (0.056) | (0.064) | | (0.042) | (0.029) |
| Ex-post Engl Diff (U-F) | 0.173* | 0.177*** | Ex-post Engl Diff (F-U) | 0.140 | 0.117** |
| | (0.076) | (0.044) | | (0.069) | (0.034) |
| Size of Majority | −0.019 | −0.059** | Size of Majority | −0.036 | 0.013 |
| | (0.026) | (0.021) | | (0.031) | (0.015) |
| Constant | 0.165 | 0.330** | Constant | 0.310* | 0.074 |
| | (0.127) | (0.097) | | (0.124) | (0.080) |
| Observations | 9,600 | 9,600 | Observations | 9,280 | 9,600 |
| R-squared | 0.210 | 0.152 | R-squared | 0.163 | 0.073 |
| Number of Subjects | 30 | 30 | Number of Subjects | 29 | 30 |

Note: OLS fixed effects regressions with punishment points and reward points as dependent variables. Punishment points (left side of the table) can take values from 0 to 7 and are used in the treatments *Punishment* and *Both*. Reward points (right side of the table) can take values from 0 to 7 and are used in the treatments *Reward* and *Both*. *Choice (Un)fair* equals 1 if the (un)fair allocation is chosen. *Outcome (Un)fair* is a dummy that equals 1 if the (un)fair outcome is implemented. *Intention (Un)kind* equals 1 if a voter votes for the (un)fair allocation while no majority was reached before. *Initiator (Un)fair* equals 1 if a voter is the initiator for the (un)fair outcome. *Pivotal (Un)fair* is an indicator that equals 1 if a voter is pivotal for the (un)fair outcome. *BF Responsibility (Un)fair* and *Ex-ante* and *Ex-post Engl Difference (U-F/F-U)* correspond to the responsibility measures explained in Sect. 3. *Size of Majority* indicates the number of majority voters and can take values from 3 to 5. Robust standard errors in parentheses clustered at the subject level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

sanction motives, which are included in this regression by other variables.[17] The

[17] The regressions in Appendix A.5 show that the responsibility measures do not contribute much on top of the other variables, and their inclusion in the regression also does not much change the coefficients of the other variables. However, whether the positional variables are included or not affects explanatory power and coefficients of the responsibility measures. Their strength has to be assessed in the individual regressions as discussed above.



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

results of the econometric comparison indicate that in sequential decisions, subjects mainly focus on the choices and the pivotal decision-maker when attributing responsibility.

Since the criterion Choice seems to be a crucial factor, we will zoom in on its role. First, we investigate what explains reward and punishment on top of Choice. So, we separately investigate what explains sanctions for fair and unfair choices. The regressions are presented in Table A26 in the appendix. On the one hand, they show that after a fair choice, there is almost no punishment and subjects do not differentiate between the different voters who voted for the fair outcome when punishing. The same is true for reward in case of an unfair choice. This result can also be seen in Fig. 3. On the other hand, the regressions show that among the voters with an unfair choice, the punishment is significantly explained by Pivotality. Again, the same is true for the reward among the voters with a fair choice. This suggests that the choice is a necessary condition for sanctioning. If the choice is fair, then there is no punishment and when the choice is unfair, there is no reward. Otherwise, more sophisticated criteria come into play.[18] Second, we explore whether subjects shift their strategies over time, especially since there are many decisions to take. For example, Choice is a relatively easy criterion and could become more pronounced towards the end of the experiment. This simpler (heuristic) strategy might come at the cost of other criteria, in particular Pivotality. The Tables A35 and A36 in the appendix show the regression tables for decisions 1-21, 22-42, and 43-64. Choice does not lose predictive power over time and Pivotality becomes even more important over time.

#### 4.1.2 Heterogeneity

While the different theoretical models and motives have shown to be important in explaining the responsibility attribution pattern on average, the explanatory power of these measures is still quite low (see Fig. 4). A potential reason for the low explanatory power of the joint analysis is heterogeneity in individual behavior. We perform finite mixture models to test if the overall data can be better explained by a mixture of different subgroups.[19] Using a bootstrap likelihood ratio test with 100 bootstrap replicates, we test if a model without subgroups (one component of coefficients) is a better fit to the data than a model with subgroups (more than one component of coefficients). In each treatment, the tests suggest that heterogeneity in individual behavior exists as the data can be better explained by a mixture of more than one subgroup ($p< 0.02$ for all treatments).

[18] In Appendix A.8, we present a hurdle model, in which we show what determines whether to sanction and what determines how much to sanction. However, this analysis does not provide new insights.

[19] We estimate a finite mixture model for each treatment (*Both* treatment separated by punishment and reward points) with the same positional variables as in Table 2. Varying the set of regressors to the full set of variables doesn't change the overall classification of components. The finite mixture models are estimated via a general linear regression using an EM-algorithm. We report the vector of coefficients for each subgroup across treatments in the Appendix A.10.



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

Although there is only a limited number of subjects per treatment, we explore the individual patterns in each treatment. We identify the optimal number of clusters in each treatment by testing the best goodness of fit. Three different punishment and three different reward patterns are present in our experiment.

Figure 5 shows the average punishment and reward points for the different voter categories used by recipients categorized within the same cluster. Considering only punishment points, 30% of the subjects in the *Punishment* treatment use no or little punishment (Cluster 1).[20] However, most of the subjects in the *Punishment* treatment punish intentional voters and especially the pivotal voter for unfair outcomes and are categorized in Cluster 2. Subjects in Cluster 3 focus on punishing unfair choices. The cluster analysis in the *Reward* treatment shows that 25% in our experiment only use little reward (Cluster 1). Subjects in Cluster 2 focus on rewarding intentional voters for fair outcomes (especially the pivotal voter), while subjects categorized by Cluster 3 show a tendency to reward fair choices.

Overall, the patterns shown in the treatments *Punishment* and *Reward* are very similar as these clusters can be described by: no or little punishment/reward, focus on the pivotal voter, and focus on choices.

The finite mixture models in the *Both* treatment are performed separately for punishment and reward points and the results are shown in Fig. 6.[21] The punishing as well as the rewarding behavior in the *Both* treatment can be described by three patterns: no punishment/reward, little sanctioning and sanctioning choices. These patterns are different than the ones observed in the *Punishment* and *Reward* treatment and suggest a less differentiated approach.

In sum, the presented exploratory cluster analysis indicates that there exist different types of responsibility attribution patterns. These results should be taken with care as there are only roughly 30 subjects per treatment and some clusters are rather small. Nonetheless, the patterns are interesting and suggest different types which are in line with different motives. We identify very similar punishment and reward patterns in our experiment. The expressed patterns are: little sanctioning, sanctioning according to intentions / pivotality and sanctioning according to choices. We find heterogeneity in individual behavior which can be an explanation of the low predictive power of the theoretical models and motives in Table 2.

After having analyzed how recipients on an aggregate and individual level attribute responsibility we complement the behavioral analysis by looking at the behavior of the voters.

[20] There are two subjects who do not use any punishment point across all decisions in the *Punishment* treatment.

[21] Note that the magnitudes of average punishment and reward points shown in Fig. 6 for each cluster in the *Both* treatment are smaller than the corresponding comparison in the *Punishment* and *Reward* treatments of Fig. 5. This can be explained by our design since in all treatments the maximum number of sanction points is limited to seven. This means that recipients in the *Both* treatment can use up to seven points for rewarding and punishing while in the other two treatments the seven points can be used for the single sanction option. We added a minimal noise of absolute magnitude $\leq 0.01$ to the reward points when performing the finite mixture model since the EM algorithm otherwise doesn't converge.



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

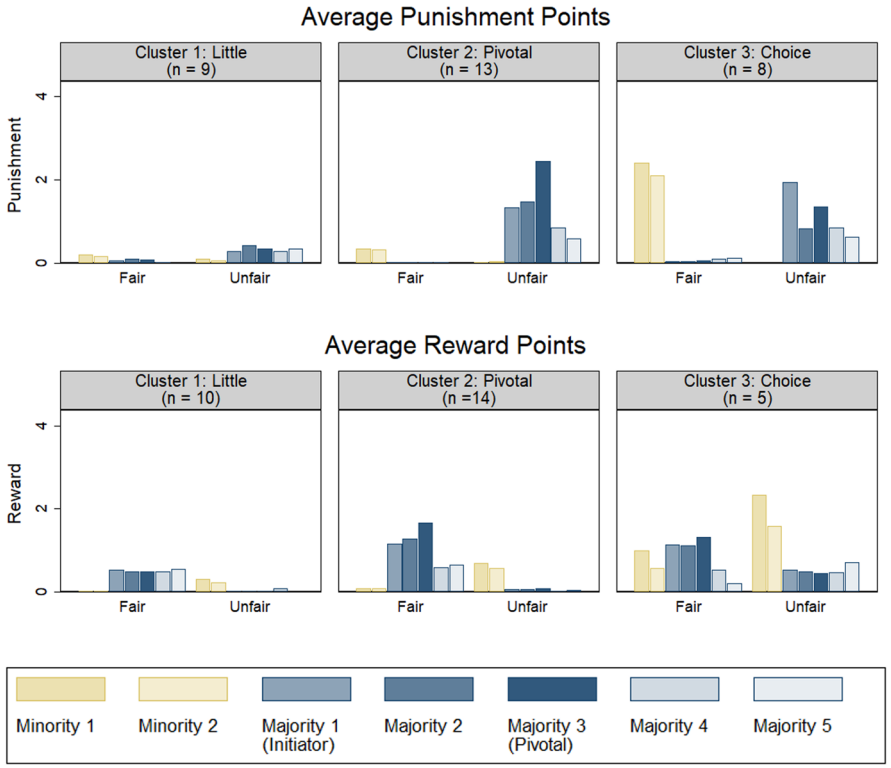



**Fig. 5** Cluster analysis: Punishment and reward patterns in *Punishment* and *Reward* treatment based on finite mixture models. Note: The figure shows the average punishment and reward points in absolute terms used in each cluster for fair and unfair outcomes across treatments. Hereby, the punishing patterns in the *Punishment* treatment are presented in the upper part of the figure, while the reward patterns in the *Reward* treatment are presented in the lower part of the figure. The number of subjects contained in each cluster per treatment are indicated in the titles of each sub-figure

### 4.2 Voting behavior

We now turn to the voting behavior and analyze how subjects vote in collective decisions under the prospect of reward and punishment. Pivotality is an important aspect in the process of responsibility attribution as we have shown in the analysis of the recipients' behavior. Accordingly, strategic voters might take this into account and prevent (favor) being the target of punishment (reward) linked to pivotality.

We examine this strategic behavior by studying the voters' behavior in potentially pivotal decisions.[22] Table 3 offers a general overview of the share of decisions in which voters choose the unfair allocation depending on their position and

[22] Potentially pivotal means that exactly two of the previous voters voted for the same allocation and that the own vote can be deterministic for the outcome. These situations can only appear for voters on positions three, four and five.



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

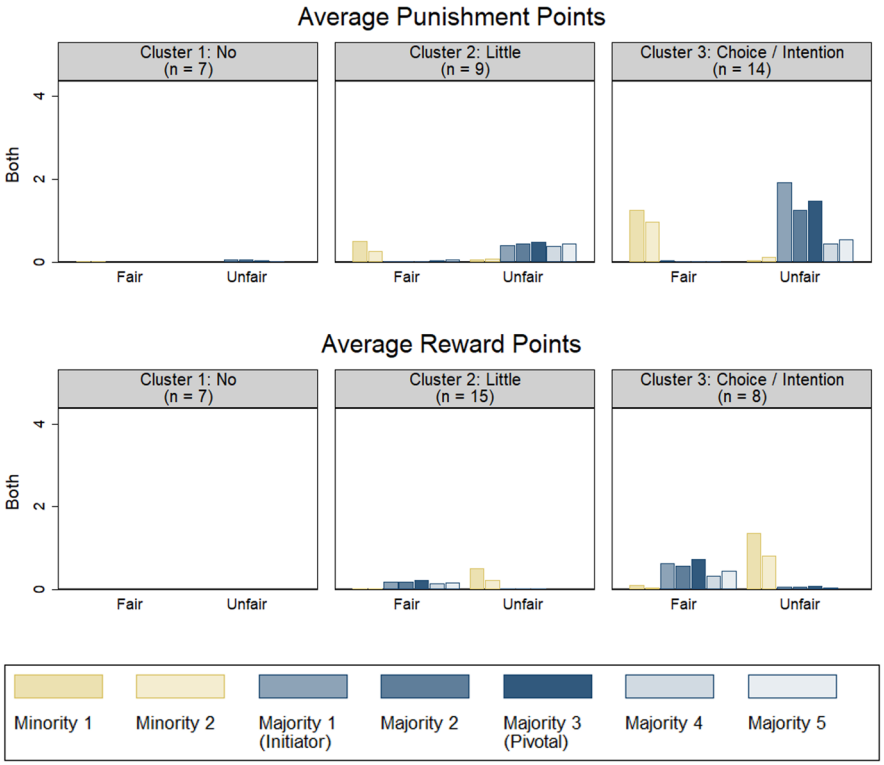



**Fig. 6** Cluster analysis: Punishment and reward patterns in the *Both* treatment based on finite mixture models. Note: The figure shows the average punishment and reward points in absolute terms used in each cluster for fair and unfair outcomes in the *Both* treatment. Hereby, the punishing patterns are presented in the upper part of the figure, while the reward patterns are presented in the lower part of the figure. The number of subjects contained in each cluster per treatment are indicated in titles of each sub-figure.

the previous votes. We separate the results by our treatments (columns 3-5) and by the decisions of the voters in the two dictator decisions (columns 6-8). Importantly, we use the dictator game to elicit the preference of the voters for the fair or unfair allocation when neither a collective decision nor punishment or reward are implemented. 63 voters show a preference for the unfair allocation, while 14 voters show a fair preference. The remaining 13 voters have mixed preferences depending on the two allocation sets we offer them. Bold sequences in column 2 indicate situations in which voters face a potentially pivotal decision.

Voters who face a potentially pivotal decision are influenced by the choices of the previous voters. Across all treatments, the share of unfair choices for potentially pivotal voters is higher when the majority of previous voters voted unfair in comparison



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

**Table 3** Voting Behavior - Share of unfair choices

| | | By treatment | | | By preference in dictator games | | |
|---|---|---|---|---|---|---|---|
| | | | | | Fair | Mixed | Unfair |
| Voter | Previous Voters | Both | Punishment | Reward | (14 Voters) | (13 Voters) | (63 Voters) |
| 1 | - | 0.45 | 0.53 | 0.43 | 0.07 | 0.58 | 0.54 |
| 2 | U | 0.55 | 0.63 | 0.40 | 0.00 | 0.69 | 0.61 |
| | F | 0.48 | 0.47 | 0.42 | 0.04 | 0.54 | 0.53 |
| 3 | **UU** | 0.40 | 0.50 | 0.48 | 0.11 | 0.62 | 0.51 |
| | **FF** | 0.35 | 0.28 | 0.40 | 0.07 | 0.46 | 0.38 |
| | Tie | 0.48 | 0.60 | 0.49 | 0.11 | 0.60 | 0.60 |
| 4 | UUU | 0.23 | 0.35 | 0.40 | 0.18 | 0.54 | 0.32 |
| | FFF | 0.02 | 0.02 | 0.12 | 0.00 | 0.00 | 0.07 |
| | **2 of 3 U** | 0.42 | 0.52 | 0.53 | 0.07 | 0.46 | 0.59 |
| | **2 of 3 F** | 0.47 | 0.44 | 0.52 | 0.13 | 0.44 | 0.56 |
| 5 | 3 or more U | 0.31 | 0.30 | 0.40 | 0.26 | 0.52 | 0.31 |
| | 3 or more F | 0.05 | 0.03 | 0.14 | 0.02 | 0.05 | 0.09 |
| | **Tie** | 0.48 | 0.54 | 0.61 | 0.11 | 0.47 | 0.65 |

Note: Bold marked sequences in column 2 indicate decisions in which voters are potentially pivotal

to a fair majority of previous votes (columns 3-5).[23] The biggest discrepancy results in the treatment *Punishment* for potentially pivotal voters on voting position three. 50% of these voters' decisions are unfair when the first two voters voted unfair, while only 28% of the decisions are unfair when the first two voters voted fair.

The sequential decision design allows voters to use strategic (non-)delegation in order to avoid (seek) pivotality. Here, we focus on the potentially pivotal voters on either position three or four. In these cases the voters can ensure being pivotal by following the majority of previous votes. But the voters can also vote against the majority of previous voters and can therefore delegate the notion of being pivotal to the next voter. Subjects showing a preference for the fair allocation in the dictator game mostly choose the fair allocation when being potentially pivotal (column 6 in Table 3). In contrast, voters expressing a preference for the unfair allocation in the dictator game often behave against their true preference in potentially pivotal situations of the collective decision (column 8). On voting positions three and four, potentially pivotal voters with an unfair preference ensure the unfair outcome by being pivotal in only 51% and 59% of all decisions respectively. In 62% and 44% of the cases where the majority of previous voters chose the fair allocation, potentially pivotal voters on positions three and four voted against their true preference and decided on being pivotal for the fair outcome. On voting position 5 where no

[23] One exception are the decisions of potentially pivotal voters in fourth position in the treatment *Both* where 47% of decisions are unfair following a fair majority, while only 42% are unfair following an unfair majority.



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

strategic (non-)delegation is possible voters followed their true preference in only 65% of the cases.

Taken together, voters showing a preference for the unfair outcome often vote against their preference when being potentially pivotal. They avoid (seek) being pivotal for unfair (fair) outcomes by strategic (non-)delegation. But surprisingly many subjects do not try to appear fair when the outcome can no longer be changed, in particular when the outcome of the decision is unfair.

This completes our behavioral results. We now turn to the processing data we collected for recipients and voters to complement the behavioral analyses.

### 4.3 Process measures

For the voters, we collected response time as a process measure. We hypothesize that response times inform us about the decision-making process, the difficulty of the decision and strategic decision-making (Konovalov and Ruff, 2022; Konovalov and Krajbich, 2019; Hausfeld et al., 2020; Spiliopoulos and Ortmann, 2018). A decision has to be considered as difficult if the voter strategically votes against the outcome preferences when being potentially pivotal. These decisions are characterized by a higher internal conflict, and should be accompanied by longer response times (Rubinstein, 2007). In Sect. 4.2, we showed that voters often decide against their true preference when being potentially pivotal. Potentially pivotal means that exactly two previous voters voted for the same allocation and that the outcome can now be determined by the respective voter. This behavior suggests an internal conflict of being potentially pivotal.

Table 4 shows how the response time of voters is affected by being potentially pivotal, the voter position and the choice of the voters. When accounting for the voter position, we find that voters take significantly more time in choosing an allocation when they are potentially pivotal (actual effect size is around 590 ms).[24] Another conflict shows when separating the voters by the true preference shown in the dictator game (last three columns of Table 4). First, being potentially pivotal still leads to higher response times for all types of voters. However, if voters vote against their true preferences, we find response times to be similarly affected. We find that fair types of voters take significantly more time when choosing the unfair allocation. In contrast, mixed and unfair types of voters spend less time when choosing the unfair allocation. Together with the results presented in Sect. 4.2, we can answer our fourth and fifth research question on how voters respond to the incentives created by responsibility attribution. Voters are aware of the responsibility that is linked to pivotality as they strategically use delegation to avoid punishment or non-delegation to gain reward even if it means that they vote against their true preference. This behavior is accompanied by a higher response time.

[24] The average log decision time over all voting decisions is 1.48 which results in an average decision time of 4.41 s. Multiplying this average decision time by the coefficient of being potentially pivotal for the whole sample results in an actual effect size of 0.59 s.



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

**Table 4** Response Time Analyses Voters

| | Dependent Variable: Log Decision Time Voters | | | | | | |
|---|---|---|---|---|---|---|---|
| | Punishment | Reward | Both | Total | Fair Type | Unfair Type | Mixed Type |
| Potentially Pivotal | 0.164*** | 0.126** | 0.114** | 0.135*** | 0.107* | 0.154*** | 0.160** |
| | (0.031) | (0.040) | (0.036) | (0.021) | (0.036) | (0.028) | (0.042) |
| Voter Position | 0.037* | 0.010 | 0.038** | 0.028*** | 0.026 | 0.021 | 0.030 |
| | (0.014) | (0.013) | (0.012) | (0.008) | (0.018) | (0.011) | (0.018) |
| Choice Unfair | −0.095 | −0.135 | −0.102 | −0.115* | 0.252* | −0.165** | −0.156* |
| | (0.057) | (0.101) | (0.076) | (0.048) | (0.098) | (0.057) | (0.070) |
| Constant | 1.346*** | 1.397*** | 1.300*** | 1.351*** | 1.259*** | 1.379*** | 1.490*** |
| | (0.073) | (0.087) | (0.078) | (0.047) | (0.084) | (0.062) | (0.103) |
| Observations | 1,860 | 1,860 | 1,860 | 5,580 | 868 | 3,906 | 806 |
| R-squared | 0.032 | 0.019 | 0.023 | 0.024 | 0.029 | 0.033 | 0.036 |
| Number of Subjects | 30 | 30 | 30 | 90 | 14 | 63 | 13 |

Note: OLS regression with Decision Time of Voters (taken in log) as dependent variable. *Potentially Pivotal* is a dummy variable indicating whether the voter faces a decision where she can be pivotal depending on her choice. *Voter Position* indicates the position the voter holds in the decision she faces and can range from 1 to 5. *Choice Unfair* is a dummy variable indicating whether the voter chose the unfair or fair option. Robust standard errors are clustered on individuals in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

For the recipients, we collected gaze data as a process measure. We tracked the gaze of participants in order to analyze whether their sanctioning behavior is reflected in their information search. The gaze data reveals interesting insights about the importance of saliency, positioning of voters and the impact of sanctioning motives on the share of fixations. We present and discuss these analyses in Section A.11 of the appendix.

## 5 Conclusion

In our study, we use reward and punishment to investigate how people attribute responsibility in decision chains. In our experiment, five voters choose sequentially between two options of how to allocate points between voters and recipients. One option is fair, the other is unfair. The recipients can reward and/or punish the voters, which we take as our measure of responsibility attribution. We test the relevance of different motives and measures and find that the actual choice, i.e., fair or unfair, plays a dominant role. It is clearly the most important determinant for the decision on whether to punish or not. Nevertheless, in line with Bartling et al. (2015), we find that the pivotal voter is assigned more punishment if the voters vote for an unfair allocation. We extend this result, showing that pivotality also matters when the voters vote for the fair allocation and the recipients can reward. Even though people have rather sophisticated responsibility attribution patterns, the conceptual models



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

of Bartling and Fischbacher (2012) and Engl (2022) explained surprisingly little in comparison to, and in particular on top of the simple and mechanistic idea that it is the deed that determines responsibility. Overall, the explanatory power of the measures is not very high. This is partly due to heterogeneity. We show, using a finite mixture model, that three classes provide a better fit. Even though the sample size is too low to make strong statements about the distribution of types, in our sample there are subjects who sanction little, subjects who focus on choice and subjects who sanction the pivotal more than the others. Our results also show that the voters are aware of how responsibility is attributed and are particularly concerned when they are pivotal, which is also visible in longer response times.

Of course, specific features of the experiment could be relevant for the outcome, for example when investigating the role of the initiator. In our experiment, the two options were chosen with similar frequency. It is possible that when an action is rarely chosen, the initiator is assigned a higher level of responsibility. This could in particular be relevant in the case of fatalities. Another feature of our design could have reduced the relevance of the pivotal voter. Voters still had to vote, even when the decision was already made. Finishing the procedure when the result is determined could make the pivotal voter even more focal. Such variants could reveal the sensitivity of responsibility attribution to the specific situation. In addition, investigating costless reward and punishment could reveal whether selfish people apply different patterns of responsibility attribution. However, one can argue that also outside of the lab, reward and punishment bear some cost and therefore it is more important to know the responsibility attribution of people who are willing to bear such cost.

The responsibility attribution in our voting game captures situations, in which people take sequential decisions or actions that jointly generate an outcome. What does this imply in the real world, for example in the case of a disaster? First, any bad action is attributed some responsibility. Second, the pivotal person, i.e., the person after whose action the disaster was unavoidable, is generally assigned the highest responsibility. In his book, Whittingham (2004) observes that often the institutional environment is an important reason for disasters. Translated to our setting, he would consider the initiator as particularly responsible. However, we find few subjects who agree with this view. Of course, there are important differences to our lab experiments. First, in such disasters it is more difficult to identify the sequence. In particular, it is not easy to identify the pivotal agent - it is difficult to find out when the disaster was no longer avoidable. Further, the different agents are less symmetric, both with respect to their contribution to the disaster and with respect to their formal responsibility. There are also many institutional details that could matter. Further research will allow investigating such variations, and our experiment provides a framework to do so.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s10683-024-09833-1.

**Funding** Open Access funding enabled and organized by Projekt DEAL.



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

## References

Albrecht, F., Kube, S., & Traxler, C. (2018). Cooperation and norm enforcement - The individual-level perspective. *Journal of Public Economics, 165*, 1–16.

Bartling, B., & Fischbacher, U. (2012). Shifting the blame: On delegation and responsibility. *Review of Economic Studies, 79*(1), 67–87.

Bartling, B., Fischbacher, U., & Schudy, S. (2015). Pivotality and responsibility attribution in sequential voting. *Journal of Public Economics, 128*, 133–139.

Besley, T. (2006). *Principled Agents?: The Political Economy of Good Government*. Oxford University Press on Demand.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A Theory of Equity, Reciprocity, and Competition. *The American Economic Review, 90*(1), 166–193.

Charness, G. (2000). Responsibility and effort in an experimental labor market. *Journal of Economic Behavior & Organization, 42*(3), 375–384.

Coffman, L. C. (2011). Intermediation Reduces Punishment (and Reward). *American Economic Journal: Microeconomics, 3*(4), 77–106.

Duch, R., Przepiorka, W., & Stevenson, R. (2015). Responsibility Attribution for Collective Decision Makers. *American Journal of Political Science, 59*(2), 372–389.

Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior, 47*(2), 268–298.

Engl, F. (2022). A Theory of Causal Responsibility Attribution. *CESifo Working Paper Series*, (9898).

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Simoudis, E., Han, J., and Fayyad, U., editors, *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, pages 226–231. AAAI Press.

Falk, A., Fehr, E., & Fischbacher, U. (2008). Testing theories of fairness-Intentions matter. *Games and Economic Behavior, 62*(1), 287–303.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics, 114*(3), 817–868.

Feinberg, J. (1970). *Doing & Deserving; Essays in the Theory of Responsibility*. Princeton University Press.

Fischbacher, U. (2007). Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171–178.

Gerstenberg, T., Lagnado, D. A., Speekenbrink, M., & Cheung, C. (2011). Rational order effects in responsibility attributions. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.*, 33(33):1715–1720.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association, 1*(1), 114–125.

Gurdal, M. Y., Miller, J. B., & Rustichini, A. (2013). Why Blame? *Journal of Political Economy, 121*(6), 1205–1247.

Hart, H. & Gardner, J. (2008). *Punishment and Responsibility: Essays in the Philosophy of Law*. Oxford University Press.

Hausfeld, J., Fischbacher, U., & Knoch, D. (2020). The value of decision-making power in social decisions. *Journal of Economic Behavior & Organization, 177*, 898–912.



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press

Iyengar, S. (1994). *Is Anyone Responsible?: How Television Frames Political Issues*. University of Chicago Press.
Konovalov, A., & Krajbich, I. (2019). Revealed strength of preference: Inference from response times. *Judgment & Decision Making, 14*(4), 381–394.
Konovalov, A., & Ruff, C. C. (2022). Enhancing models of social and strategic decision making with process tracing and neural data. *Wiley Interdisciplinary Reviews: Cognitive Science, 13*(1), e1559.
Krupka, E. L., & Weber, R. A. (2013). Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary? *Journal of the European Economic Association, 11*(3), 495–524.
Leibbrandt, A. & López Pérez, R. (2011). Individual heterogeneity in punishment and reward. *Universidad Autónoma de Madrid. Department of Economic Analysis. Working Papers in Economic Theory*, No. 2011/01.
Oexl, R., & Grossman, Z. J. (2013). Shifting the blame to a powerless intermediary. *Experimental Economics, 16*(3), 306–312.
Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *The American Economic Review, 83*(5), 1281–1302.
Ross, L. & Nisbett, R. E. (1991). *The Person and the Situation: Perspectives of Social Psychology*. New York (N.Y.) : McGraw-Hill Book Company.
Rubinstein, A. (2007). Instinctive and cognitive reasoning: A study of response times. *Economic Journal, 117*(523), 1243–1259.
Shaver, K. G. (1985). The Attribution of Causality. In *The Attribution of Blame*, pages 35–62. New York: Springer.
Spiliopoulos, L., & Ortmann, A. (2018). The BCD of response time analysis in experimental economics. *Experimental Economics, 21*(2), 383–433.
Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York: Guilford Press.
Whittingham, R. (2004). *Design errors, Why Human Error Causes Accidents*. London: In The Blame Machine Routledge.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

**Deepti Bhatia[1,2] · Urs Fischbacher[1,2,3] · Jan Hausfeld[4,5] · Regina Stumpf[1,2]**

✉ Regina Stumpf
regina.stumpf@uni-konstanz.de

Deepti Bhatia
deepti.bhatia@uni-konstanz.de

Urs Fischbacher
urs.fischbacher@uni-konstanz.de

Jan Hausfeld
j.hausfeld@uva.nl

[1] Department of Economics, University of Konstanz, Konstanz, Germany

[2] Thurgau Institute of Economics, Kreuzlingen, Switzerland

[3] CESifo, Munich, Germany

[4] CREED and Amsterdam School of Economics, University of Amsterdam, Amsterdam, The Netherlands

[5] Tinbergen Institute, Amsterdam, The Netherlands



https://doi.org/10.1007/s10683-024-09833-1 Published online by Cambridge University Press