Helmut-Schmidt-University Hamburg

Faculty of Economics and Social Sciences

An Experimental Study on Responsibility Attribution in Sequential Decision-Making

Dissertation submitted for the academic degree of Doctor rerum politicarum (Dr. rer. pol.)

Supervisor:
Prof. Dr. Stefan Traub
Chair of Behavioral Economics
Helmut-Schmidt-University Hamburg
Holstenhofweg 85
22043 Hamburg

Janina Kraus
February 2023

Acknowledgements

Between the beginning of this dissertation in May 2018 and its end, almost five exciting, eventful, and wonderful years of ups and downs have passed. At this point, I would like to express my special thanks to the following people, without whose support the completion of this dissertation would not have been possible.

First and foremost, I would like to thank my advisor, Professor Dr. Stefan Traub. I have always found the numerous conversations on an intellectual and personal level to be stimulating and encouraging. I am very grateful for his scientific and moral support during my time as a doctoral student. I would like to thank Prof. Dr. Dierk Herzer for his kind supervision as second examiner, Prof. Dr. Michael Berlemann and Prof. Dr. Christina Schäfer for their participation in the doctoral committee.

Furthermore, I would like to thank my former colleagues at the Helmut-Schmidt-University Hamburg and my current colleagues at the TU Clausthal. In particular, the constructive exchange and regular professional discussions have always been of great help and positive encouragement to me. Thank you for the great collegial and friendly atmosphere in the office and your understanding in stressful times. I would especially like to thank Hakan for his linguistic support.

I would like to thank my friends, especially Christin, Jenny, Judith, Lea, Lisa, and Marleen, for their never-ending (moral) support and patience, especially during the last months. They gave me the strength and courage to write and complete my dissertation. Finally, my special thanks go to my parents who have always believed in me and to whom I dedicate this dissertation.

Contents

List of Tables

List of Figures

Acronyms

AER	American Economic Review	29
AIC	Akaike's Information Criterion	13, 15, 61, 63, 67
BIC	Bayesian Information Criterion	13
QJE	Quarterly Journal of Economics	29
RPP	"Reproductibility Project: Psychology"	29, 30
SCT	Status Characteristics Theory	18
SES	socioeconomic status	18

1. Introduction

"Perceptions of responsibility are a pervasive component of everyday live."

(Engl 2018, p. 1)

People are constantly making decisions, the results of which often depend not only on individual choices but also on the choices of others and can have consequences for others who are not part of the decision process itself. For example, strategy decisions in companies are often made by a board of directors rather than by individual managers, and affect (all) employees. Policies for the people are decided by congresses, not by individual politicians. In families, partners often decide together and not individually how to spend the household money for the whole family (among others, see e.g. Katzenbach and Smith 1994; Kugler et al. 2012). If such decision-making processes are transparent, i.e., if individual decisions are observable (Bartling et al. 2015; Fischbacher and Schudy 2020), those affected can associate who decides for or against them.

The pervasiveness of decisions is clear. However, there are still many unknown factors regarding the judgment of affected parties, especially how each individual decision-maker is held accountable for his decisions and the resulting outcome. The experiment by Bartling et al. (2015), from which our experimental design is adapted, uses a sequential dictator game with punishment and identifies the pivotal decision-maker as the focal point for attributing responsibility. Duch et al. (2014) shed light on the prominent position of the proposer, while a veto player is not held responsible for letting the proposal pass and thus doing nothing. However, it is still an open question as to how these aspects combined may affect the attribution of responsibility. More precisely, combining the sequential dictator game of Bartling et al. (2015) with different (status) positions and the possibility to passively implement decisions has never been done before and may provide a more nuanced picture of how individual responsibility is attributed.

Thus, the focus of this study is on group decisions that have distributional consequences for others not involved in the decision itself. Specifically, we are interested in how these affected individuals evaluate the responsibility of the decision-maker for the individual choices and the outcome resulting from group decisions. Therefore, we use a sequential group decision with different roles and a preselected option that can be passively implemented. In particular, we are interested in the interaction between (1) the (legitimate) procedure by which status

positions are assigned and (2) the way in which decisions are implemented, either actively or passively, on the attribution of responsibility to each individual decision-maker.

As in the experiment by Bartling et al. (2015), we use assigned punishment points as a measure of responsibility. Thus, more punishment is equated with more responsibility. We hypothesize that the active (passive) implementation of decisions and the legitimacy of status differences lead to a stronger (weaker) correlation between responsibility and punishment. Consequently, it is not necessarily responsibility that changes between treatments, but rather the punishment for a given responsibility of a particular decision-maker. More specifically, if a decision-maker chooses an unpleasant allocation but has a legitimate position or does not actively decide, a recipient may be more likely to accept responsibility for an unpleasant decision and less likely to punish. To simplify, we use assigning responsibility as a synonym for assigning punishment.

Thus, this study aims to identify (additional) relevant factors that determine the attribution of responsibility by affected parties. Therefore, we focus on responsibility attribution in a transparent and sequential group decision by using the experimental design of Bartling et al. (2015) and extending it to two dimensions:

> (1) **Default Option:** either an equal or unequal allocation is already preselected and can be implemented passively.
> (2) **Status Differences:** using two group-building mechanisms that differ in their assessed legitimacy, we induce status differences and assign participants to their roles as either decision-makers or recipients.

The first extension focuses on the default effect. Specifically, we examine whether decision-makers are more responsible when the unequal allocation is actively implemented by an act of commission, given that the preselected option is equal. Alternatively, implementing an unequal allocation through an act of omission may absolve decision-makers of responsibility. Here, three concepts are distinguished: omission, default options, and status quo. The default is the preselected option (Gärtner and Sandberg 2017, p. 2) that may be implemented by an act of omission, thus doing nothing (Shultz et al. 1981, p. 241) while the last one refers to an alternative that has been implemented in the past (Samuelson and Zeckhauser 1988, pp. 7-8).

Defaults are relevant in everyday life and often have far-reaching implications. Examples include the recent discussion of mandatory vaccination against Covid19 (Serra-Garcia and

Szech 2022), retirement savings (McKenzie et al. 2006; Carroll et al. 2009), and organ donation (Johnson and Goldstein 2003; Gimbel et al. 2003). Defaults are closely related to acts of omission, since doing nothing usually implements the default option (Gärtner and Sandberg 2017, p. 2). In moral philosophy and psychology, there are many examples of evaluating who is responsible for acts of omission, since the causal relationship is less clear (Vaal 1996, p. 169-170). For example, the famous 'trolley problem', where throwing a switch can save five railroad workers by killing one person, while the default option is not to throw the switch, so five people would be killed (Bartels 2008, p. 383). In hypothetical situations like these, the question is whether a person is responsible for an active switch that saves five lives while causing the death of one. Alternatively, is he more (less) responsible for actively deciding to throw the switch than for failing to prevent it from being thrown?

In this respect, previous experimental results are quite clear. Unpleasant or harmful decisions are more often implemented by acts of omission than by acts of commission (among others, see e.g. Spranca et al. 1991; Vaal 1996). Acts of omission and choosing the default are judged less negatively than commissions, although the outcome may be the same. Consequently, harmful or unpleasant omissions and their resulting outcomes are less blameworthy than commissions because the causal role and intentions of the decision-maker are less clear (among others, see e.g. Ritov and Baron 1992, p. 50; Vaal 1996, pp. 169-170; DeScioli et al. 2011a, pp. 206-213). Given our design, passive implementation of an unequal allocation (as it is preselected) is expected to result in less punishment than actively choosing the same allocation.

The second extension is concerned with status differences between participants, or rather who has the right to decide and allocate the endowment. In most experimental settings, roles are assigned randomly, and even the random procedures vary, from coin tosses (Hoffman and Spitzer 1985) to color chips (Butler 2014) to a computer randomly assigning roles (Bartling et al. 2015). Since these procedures may not necessarily be considered fair or legitimate, a superior role (such as the decision-maker) may not be accepted by the inferior roles (the recipients), thus influencing the assigned responsibility and leading to more (less) punishment. In general, a higher status is more likely to be accepted if it results from special expertise, a central role, or an assigned task that influences one's sense of responsibility and the assignment of responsibility by others (Weiss and Fershtman 1998, p. 802; Ball et al. 2001, pp. 161-162).

Whether these status differences are accepted, however, strongly depends on their empirical legitimacy, i.e. the voluntary approval of the people (Hinsch 2008). In this context,

legitimacy is a widespread and important phenomenon, implying that something is voluntarily accepted and seen as right, leading to stable (status) structures (Zelditch 2001, p. 40). Whenever resources are unequally distributed between groups or individuals, justifying these inequalities (e.g. through a legitimate mechanism) may increase their acceptance (Cremer and Dijke 2009, pp. 114-115). However, advantaged and disadvantaged people must see the mechanism as legitimate for it to be effective (Olson and Hafer 2001, p. 157). Consequently, for our setting, this implies that if a legitimate group-building mechanism induces status differences, acceptance rates of inequalities should be higher so that those affected are less likely to punish.

However, it is still an open question as to what constitutes a legitimate mechanism for implementing status differences and assigning roles in laboratory experiments. In a pretest, we evaluate four mechanisms according to their legitimacy to assign different status positions. Specifically, we use the Raven's Matrices as a measure of cognitive ability (Putterman et al. 2011; Paetzel and Sausgruber 2018) and the final grade as a past-acquired characteristic that is unrelated to the decision situation itself but provides information about intelligence and knowledge (Wentzel 1991; Camara 2005). Third, the slider task developed by Gill and Prowse (2019) is a typical example of a real-effort task that is easy to understand and does not require knowledge. Finally, since a random procedure is quite common in laboratory experiments and is used in Bartling et al.'s (2015) experiment, we include a random computerized role assignment. We have already described the four mechanisms according to our expected legitimacy assessment, although the legitimacy of a random procedure can be evaluated in two ways. On the one hand, it is completely arbitrary and may therefore be perceived as the least legitimate mechanism. On the other hand, participants show a preference for randomization (among others, see e.g. Agranov and Ortoleva 2017; Cettolin and Riedl 2019), so that the reverse case, where a random procedure is rated as (most) legitimate, is also possible. As an unexpected result, the pretest actually finds that randomly assigned roles are rated as the most legitimate, while the final grade is the least legitimate mechanism.[1] Consequently, we use these two to induce status differences in our experiment.

Furthermore, to our knowledge, the interaction of these status differences with a default option implemented by an act of omission has not been investigated. Thus, our study aims to fill this research gap by investigating two relevant processes: (1) the group formation process and (2) the decision making process. The group-building process induces (legitimate) status differences among participants by assigning them the role of either 'decision-maker' or 'recipient'. Since the mechanism that induces these status differences is either more or less

[1] We will address this puzzle further in Section 6.1.4.

legitimate, we expect this to affect the second process, the allocation decision.

As mentioned above, the main experimental design is adapted from Bartling et al. (2015). It consists of a modified dictator game with second-party punishment, in which groups of six are formed. Participants are equally divided between decision-makers and recipients, who are determined by either a more legitimate (= random) or less legitimate (= final grade) group-building mechanism. Three decision-makers decide sequentially by majority vote on an allocation for their entire group. The allocation is either equal or unequal, with the latter favoring the decision-makers. Since the total payoff for the group is the same, welfare maximization considerations are irrelevant. One of the two allocations is already preselected and will be implemented automatically after 30 seconds if no active choice is made. In a second step, the three recipients can punish each decision-maker individually by assigning punishment points that serve as a measure of responsibility.[2]

This setting involves a dyadic relationship between two positions that can be viewed from two different angles: (1) What determines the decision of the decision-makers? (2) What determines the reaction of the recipients? In this study, we take the second perspective and analyze the punishment behavior of the recipients. Although the first perspective also provides interesting points of reference, we focus on the second perspective. We believe that the interaction of status differences with a default option seems to be more meaningful and unpredictable for the attribution of responsibility. Nevertheless, we will provide a brief overview of decision-makers' choices.

The main result of Bartling et al. (2015), that the pivotal decision-maker is punished more for an unequal outcome than the other decision-makers, cannot be replicated in this study. Instead, choosing unequal (regardless of the outcome) and being the first person to make an unequal choice leads to significantly more punishment points. Interestingly, we do not find significant treatment differences even though the random group-building mechanism is considered more legitimate, so the group-building mechanism does not seem to affect the punishment decision. However, the preselected default option does have an effect. In some situations, choosing the unequal allocation leads to more severe punishment when the equal allocation is preselected. This suggests that more responsibility is assigned when the choice is actively implemented. Possible explanations for this unexpected behavior and further details are discussed in the second part of this study.

In summary, this study focuses on group decisions with distributional consequences for others, while we are interested in how those affected assign responsibility, in terms of punish-

[2]Using punishment as a measure of responsibility is common in laboratory experiments (among others, see e.g. Bartling and Fischbacher 2012; Bartling et al. 2015).

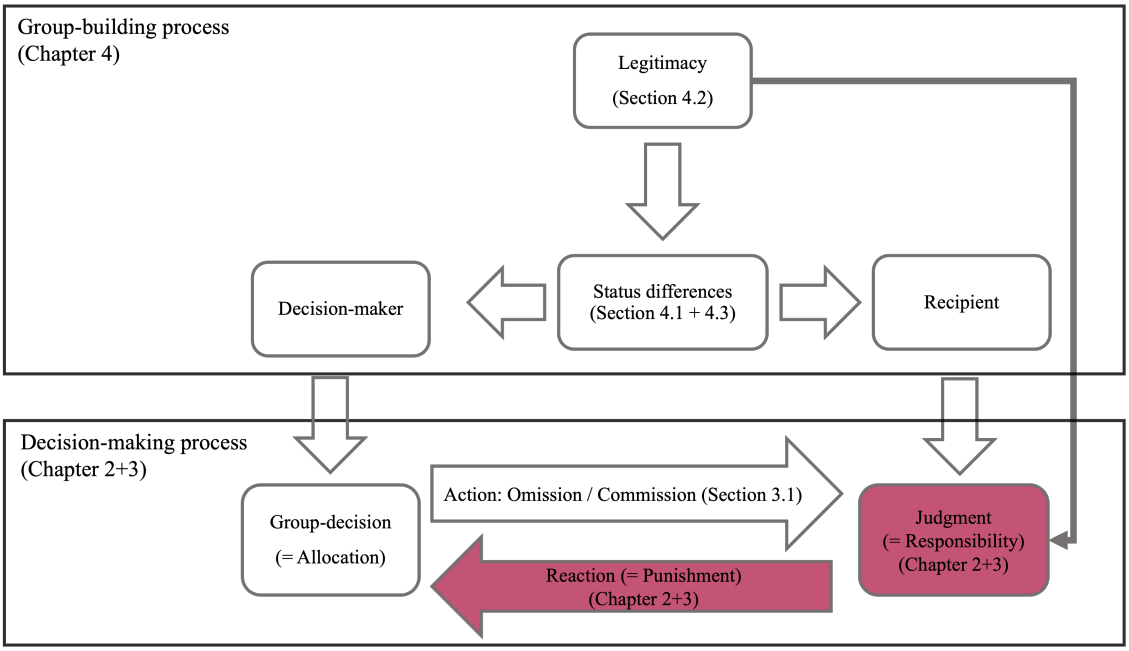


Figure 1.1.: Main Structure

ment points, to the individual decision-makers. In particular, we are interested in the process by which decision-makers and recipients are assigned to their roles and how decisions are implemented, either actively or passively, as well as the interaction between both aspects and their effect on punishment behavior. In line with our research agenda, we begin in the first part with a review of the theoretical and experimental literature. In the second part, we discuss the pretest and our laboratory experiment. Figure 1.1 visualizes the main structure of our study, with responsibility attribution and punishment reaction highlighted in red. In addition, the chapters and sections presenting the corresponding theoretical and empirical literature are indicated for structural orientation.

In Chapter 2, we focus on responsibility attribution in groups. Therefore, we start with a theoretical overview of definitions and theories, mostly from a psychological point of view (Section 2.1), followed by experimental evidence on responsibility attribution regarding decision-makers' choices and, most importantly, recipients' reactions (Section 2.2). The last section (2.3) sheds light on a political dimension of responsibility attribution.

With the second chapter laying the groundwork for how responsibility is (could be) attributed, Chapter 3 aims to complement this by focusing on the influence of the situational context from an economic perspective. Therefore, we start with the way decisions are implemented, either actively or passively, and provide theoretical and empirical evidence on omissions and defaults as well as possible explanations for their choice (Section 3.1). We

then turn to social norms (Section 3.2) and preferences (Section 3.3). Individual behavior is often influenced by appropriate norms depending on the situational context (Fershtman et al. 2012), while social and distributional preferences specify (subjective or objective) criteria for how allocations should be distributed (Croson and Konow 2009). Finally, Section 3.4 is already a transition to the next chapter because it deals with procedural fairness, i.e. the relevance of fair procedures from an economic point of view.

Chapter 4 specifies the group-building process, so how participants are assigned to different roles or positions, and the importance of legitimacy in this context. The group-building mechanism that leads to status differences must be accepted in order to be effective (Weiss and Fershtman 1998; Ridgeway 2001). One way to achieve stable status differences is through legitimate procedures that lead to greater acceptance of (unpleasant) decisions and payoff inequalities (among others, see e.g. Olson and Hafer 2001; Levin et al. 2002). Therefore, Chapter 4 provides a theoretical overview of the concept of status (Section 4.1), briefly reviewing its definitions and considerations in psychology and economics. Second, it provides a theoretical overview of legitimacy (Section 4.2) by presenting its roots and implications for status differences. The chapter concludes in Section 4.3 with experimental evidence from related studies and an overview of the group-building mechanisms typically used in laboratory experiments to induce status differences.

Chapter 5 is devoted to bringing together the theoretical and empirical evidence from the first part by providing a synthesis of what factors should (could) influence the attribution of responsibility in our setting.

The second part of our study begins in Chapter 6 with the presentation of the pretest (Section 6.1). The idea is to investigate which of the four mechanisms is rated as most (least) legitimate, as well as a brief discussion of the resulting implications. Since a random procedure is evaluated as the most legitimate, while the final grade is the least legitimate mechanism, we provide explanations for this behavior. Then, in Sections 6.2 and 6.3, we discuss our experimental design in detail and describe how we conducted the experiment at the University of Hamburg. Since our experimental design is adapted from Bartling et al. (2015), who use the strategy method while we implement the direct response method, we provide a methodological overview of elicitation procedures in laboratory experiments (Section 6.4.1) and replications in experimental economics (Section 6.4.2). Finally, Section 6.5 specifies our hypotheses based on Part 1, the synthesis in Chapter 5, and the pretest.

Then, in Chapter 7, we come to the results of our study and present how people actually assign individual responsibility. Therefore, we start with a description of our sample (Section 7.1) and present the actual choices of decision-makers as well as the expected choices of

recipients (Section 7.2). Then, in Section 7.3, we compare the rating behavior of the pretest with the experiment and focus on the deservingness of the role. Section 7.4 is the main part of this chapter and focuses on the punishment behavior of the recipients. After classifying them into different categories (Section 7.4.1), we present an overview of punishment behavior (Section 7.4.2), followed by probit and tobit regressions (Sections 7.4.3 and 7.4.4). We then present further analyses in Section 7.4.5, including robustness checks, compare our results with those of the Bartling et al. (2015) experiment (Section 7.4.6), and conclude with a comparison between actual and expected punishment (Section 7.4.7). The final chapter, Chapter 8, concludes by summarizing and discussing our study and results.

Part I.

Responsibility Attribution, Default-Effect and Status Differences - Theory and Evidence

2. Responsibility Attribution in Groups

"A fundamental assumption across all attribution theory models is that individuals have a need to infer causes and to assign responsibility for why outcomes occur. The motivation to assign responsibility is strongest when an event is surprising or unexpected."

(Blount 1995, p. 132)

The concept of responsibility and who is held responsible is prevalent in many everyday situations. Because of its importance, it is examined in many disciplines, from economics and psychology to philosophy and the legal field. Holding people responsible for ethical, moral, or legal transgressions and sanctioning these violations is crucial for social control (Schlenker et al. 1994, p. 632; Alicke 2000, p. 556). In the legal domain, for example, many researchers identify causation as a prerequisite for responsibility, which is necessary to be liable. Without causation, there should be no legal liability and, thus, no sanctions (among others, see e.g. Hart and Honoré 1985; Epstein 1973). In psychology, two dimensions are commonly distinguished: on the one hand, causal responsibility for an unpleasant outcome (= harm) and, on the other hand, the intention to harm, with the latter being closely related to various mental states, such as the beliefs or desires of the actor. Cushman (2008) specifies that the assignment of responsibility is strongly linked to the actual harm caused by the decision-maker (p. 354).

Numerous theoretical and empirical studies have addressed the question of how responsibility should be assigned and what exactly the relationship between causation and responsibility is (among others, see e.g. Shultz et al. 1981; Darley and Shultz 1990). In examining the attribution of responsibility two prominent lines of research exist. First, some scholars focus on the causality of the actor, so how the decision-maker's causal role determines the assigned responsibility and what (potential) factors mitigate his responsibility (among others, see e.g. Mikula 2003; Malle et al. 2014). The second line of research focuses on the action's causality and the impact it has in realizing the (unpleasant) outcome (among others, see e.g. Spellman 1997; Chockler and Halpern 2004). Thus, Section 2.1 starts with a definition of causality and then reviews these theoretical lines of research.

The present study is concerned with the economic perspective, in which responsibility attribution is examined using laboratory experiments. Therefore, in Section 2.2 we assess the most relevant experiments of individually assigned responsibility for collective decisions.

Some notable experimental studies focus on the assigned responsibility in sequential voting settings (Duch and Stevenson 2014; Bartling et al. 2015; Anselm et al. 2022), examine the role of proposal power (Duch et al. 2014) and agenda power (Fischbacher and Schudy 2020).

Since group or committee decisions play an important role in the political sphere, whether in the parliament or within a political party, an extensive literature deals with the attribution of responsibility in political science. Of particular interest is the evaluation of a political institution, such as the parliament or the incumbent party, and the influence of voting rules. Section 2.3 aims to provide a short overview of the most relevant theoretical and empirical findings in political sciences to specify relevant aspects for the attribution of responsibility. More specifically, the political perspective identifies the pivotal voter (Shapley and Shubik 1954) and the value of incumbency as relevant clues for the attribution of responsibility. Additionally, the voter's economic performance (Kinder and Kiewiet 1981) and sophistication (Gomez and Wilson 2003) are highlighted and classified in a broader context.

2.1. Theories of Responsibility Attribution

This section provides an overview of the main models and theories dealing with the attribution of responsibility and moral judgments. It highlights the relevant dimensions and factors that have an impact on responsibility. Of course, far from being mutually exclusive, they sometimes overlap, while other considerations may also affect the attribution of responsibility, which cannot be all presented here. Generally, two dimensions of causality can be distinguished: (1) causality of the actor, (2) causality of the action, and (3) their combination. Before turning to present each of these dimensions at a time, we start with a definition of causality. Nearly all models have been tested empirically (mainly using vignette studies) and confirmed in these parochial design contexts. However, they have not been tested against each other. Thus, no conclusion about the overall effect and the best-predicted model of responsibility assignment is possible.

2.1.1. Definition

The relationship between actions and the resulting consequences is almost always relevant when thinking about decisions. More specifically, does a change in Y lead to a change in X, or is the relationship inverse, so that X changes Y? Thus, to predict the consequences of actions, it is necessary to determine the (causal) relationship between them – that is, the direction of action (Orcutt 1952, pp. 305-306). A correlation between two variables X and

Y does not imply a cause-effect relationship since correlations are usually symmetrical. In contrast, causality is defined as an asymmetric concept, i.e., a relationship in one direction (X causes Y). This concept is based on the idea that the cause is not downstream of the induced effect (Orcutt 1952, p. 305). A causal asymmetry implies "that one can use causes to manipulate their effects, but not vice versa" (Hausman 1986, p. 143).

According to Hume (1948), a causal priority, in the sense that a cause precedes the corresponding effect, is the central aspect of the asymmetric relationship between cause and effect. In other words, the statement that Y is a cause of X is valid if the two are causally related, while X must causally precede Y (Mackie 1974, p. 51). Thus, if Y is a cause of X, this means that a change in Y leads to a change in X that would not have occurred if Y had not changed. Of course, X could also change because of some other unknown factor (Orcutt 1952, pp. 305-306). For example, a choice in a dictator game (Y) causes the outcome (X) and a dictator chooses unequally, so that the unequal outcome is implemented. Then a different choice (equal instead of unequal) leads to a different outcome (here the equal outcome) that would not have occurred if the choice (Y) had not been changed from unequal to equal. Consequently, we define causality in this study as follows.

Definition 1: Causality

Causality is an asymmetric relationship between two variables in which one variable causes and precedes the other. A change in the cause entails a change in the effect, but not vice versa.

There are many statistical tests and procedures to identify a causal relationship. For example, in economics, the concept of Granger causality is often used in the analysis of time series data to determine the relationship between two variables when the effective direction is *a-priori* unknown (Granger 1969, 1988). As we do not have time-series data in our experimental setting, this aspect is not further discussed here.

2.1.2. Causality of the Actor

The presented models in this section are classified as decision-stage models consisting of a stage-like framework.[3] They are normative and specify how people should assign respon-

[3]This expression is based on the classification of Langenhoff et al. (2021), who classified the models in decision-stage models and computational models. However, the idea of a stage-like framework has its roots in the early approaches of psychologists and philosophers such as Kohlberg (1969).

sibility, while they do not convey explanatory factors for this assessment. A further stage is only reached when all conditions of the previous stage are met. Therefore, these models start with a causal analysis of the decision situation and then address further relevant factors for assigning responsibility. The evaluation criteria at each stage are quite different and range from objective factors (like causation) to subjective factors, such as foresight and intention (Alicke 2000, p. 557; Langenhoff et al. 2021, p. 3). In the following, the most relevant decision-stage models are presented.

Early Approaches

One of the earliest concepts in the field of responsibility attribution is the work of Piaget (1932) and Heider (1958). Piaget (1932) begins by focusing not only on the outcome of an action, but on the intention behind that action. Heider (1958) never defines a clear concept of attribution, but uses five different dimensions to define how responsibility is assigned. These five dimensions are association, commission, foreseeability, intentionality, and extenuating circumstances. Thus, the decision-maker must be associated or related to the outcome, the outcome must be the result of a committed act (active choice), and the outcome must be foreseen. Moreover, a person is responsible only for intended results; if extenuating circumstances exist, they may reduce the personal responsibility (pp. 112-114). This early classification already contains essential aspects, such as causality, intention, and commissions that are relevant for attributing responsibility and will be further addressed throughout our study.

Based on the notion of Heider (1958), Shaver (1985) proposes an approach that is one of the most comprehensive approaches on responsibility. It distinguishes five different prescriptive dimensions that a person might use to attribute responsibility. These dimensions are as follows:

- *Causality* means that the decision-maker is a (direct) cause of the outcome.
- *Knowledge* implies the decision-maker being aware of the consequences.
- The degree of the intention behind the action is considered *intentionality*.
- Strong (external) forces, such as *coercion*, mitigate the responsibility for an intentional action.
- The decision-maker has to be aware of the *(moral) wrongfulness* of his action.

Shaver (1985) chooses to sequence these dimensions as people might proceed in assigning responsibility (pp. 101-113). Thus, causality is required for assigning responsibility, as is the awareness of the resulting consequences and the intention behind the action. Assuming the decision-maker is not causally responsible for the outcome or is unaware of the (potentially harmful) consequences of his actions, he is not considered responsible. Moreover, the use of coercion by an outsider may mitigate the assigned responsibility. Thus, someone may be the cause of an outcome by having control over the implementation, but is not held accountable due to mitigating circumstances – for example, a student who does not try hard to get a better grade because he decides (voluntarily) to support his ill mother at home (Weiner 1993, p. 959).

These early approaches and examples already show the *multidimensional construct* of responsibility attribution, where several dimensions and factors are relevant. The multidimensionality is further analyzed in Gailey and Falk's (2008) study. They use two vignette studies (with more than 600 undergraduate students) to empirically test the five dimensions of responsibility attribution developed by Shaver (1985). Participants are required to rate 20 questions corresponding to the five dimensions on a scale of 0 to 10. Gailey and Falk (2008) show that the theoretical model works quite well as a multidimensional concept. People focus on four of the five dimensions when attributing responsibility, more specifically, on the cause, knowledge, and intention of the actor, as well as whether the actor is aware of the wrongfulness of his action (p. 672).

Summary 2.1: Early Approaches on the Attribution of Responsibility

People consider different factors when assigning responsibility. Consequently, the attribution of responsibility may vary depending on the situational context and between people if they have different contextual beliefs.

However, these early models do not focus on the precise relationship or causality *between* the various factors and dimensions. Therefore, the following sections focus on more advanced models that show a (causal) relationship between different aspects of responsibility attribution, thereby influencing the decision to punish.

Further Relevant Models and Concepts

One of the first attempts to examine the relationship between causation and (moral) responsibility in more detail is made by Shultz et al. (1981). Their theoretical model, visualized

in Figure 2.1, illustrates the relationship between causation, moral responsibility, and punishment. It defines responsibility as the (moral) evaluation of the decision-maker's action. Punishment is the consequence of causing harm, which occurs when responsibility for the harm is assigned (pp. 239-241).

Figure 2.1.: Relationship between Causality, Moral Responsibility and Punishment

Shultz et al. (1981) specify that "a cause concerns the issue of whether or not the harm was produced by the protagonist, moral responsibility refers to the extent to which the protagonist is worthy of blame, punishment deals with the recommended consequences of the protagonist" (p. 242). Thus, to assess a person's responsibility, the decision-maker's causal role is a predisposition, which, in turn, is a predisposition of punishment. If the decision-maker is not seen as the cause and as (moral) responsible for the outcome, punishment considerations do not arise (Shultz et al. 1981, p. 250).

In this context, an action is defined as the cause of an outcome if another action changes the outcome (Wells and Gavanski 1989, p. 161). Following the legal argumentation, a necessary condition means that the decision-maker is worthy of harm if the unpleasant outcome only occurs because of his action. In contrast, a sufficient condition states that a decision-maker is the cause of harm if his action deviates from the appropriate norms and violates the normative standards of the situation. Shultz et al. (1981) summarize this as follows: "The essential logical properties of necessary conditions can be represented as *q only if p*, and those of sufficient conditions as if *p then q*" (p. 239). In this respect, p stands for the decision-maker's action and q for the resulting unpleasant outcome. If the decision-maker's behavior is a necessary condition for the harm, he is also morally responsible for his action and suffers a higher punishment probability. In contrast, actions that are a sufficient condition do not influence the judgment of causation (Shultz et al. 1981, p. 245).

An extension of the previously presented approach is the Shultz-Schleifer Model (see Figure 2.2). It additionally addresses the role of blame and distinguishes two forms of responsibility. One is moral responsibility, as described above, and the other is vicarious or passive responsibility. According to the Shultz-Schleifer Model, there are situations in which blame is assigned without the decision-maker's causality. For example, parents are vicariously liable for their younger children. As can be seen in Figure 2.2, moral responsibility presupposes causation, blame presupposes moral responsibility, and punishment presupposes blame (Darley and Shultz 1990, p. 531). Consequently, the decision-maker is not morally responsible if

the harm is accidental (i.e., caused without any intention to harm anyone), involuntary (i.e., caused under coercion), or if the harm was not foreseen.

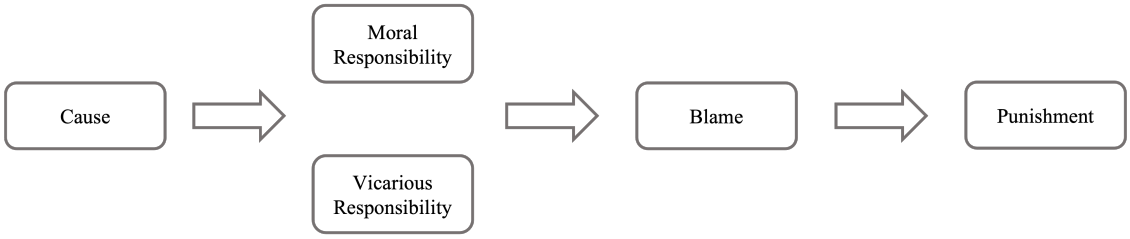


Figure 2.2.: The Shultz-Schleifer Model

Three other models, extending the relationship between causation, responsibility, and punishment, are the triangle model by Schlenker et al. (1994), the culpable control model by Mark Alicke (2000), and the Path Model of Blame by Malle et al. (2014). Since each of them highlights at least one interesting aspect, they are (briefly) presented here.

The triangle model by Schlenker et al. (1994) specifies parts of the personal identity as determinant for the attribution of responsibility. Three main criteria (or links arranged in a triangle) are defined. (1) The first criterion concerns the relationship between the behavior and existing cultural prescriptions. These rules contain explicit or implicit information about the goals and norms for behavior and the appropriate way to achieve them, and are used to guide and evaluate the behavior. (2) The second criterion analyzes whether the actor acts according to these prescriptions. (3) The third one considers whether the actor acts voluntarily and intentionally. Thus, the actor seems to be connected to the event and the relevant prescriptions, especially by having personal control over the event (Schlenker et al. 1994, pp. 634-635).

The culpable control model by Alicke (2000), which is complementary to the triangle model of Schlenker et al. (1994), identifies conditions that reduce or increase the assigned blame. The model identifies three dimensions of personal control relevant for the blame assignment, which are (1) the mental state, i.e., a person's desires, plans, or motives; (2) the behavior itself, i.e., the actual action, which can be performed either actively or passively; (3) the resulting consequences, i.e., the outcome (p. 557). Therefore, the model cites intention, causation, and (or) foresight as aspects of personal control that can increase or decrease the assigned blame. Unlike other models, personal control can be downgraded in the culpable control model, so that different degrees of blame can be assigned (Alicke 2000, pp. 559-570).

An interesting aspect is highlighted by Cushman (2008). Using vignette studies, he examines the distinction between judgments of wrongness and blame. Here, the former is

determined by the decision-maker's beliefs and desires and the latter by the actual consequences of an action. For the wrongness of an action, the intention of the decision-maker is far more important than the consequences of his action. This supports the idea that a decision-maker's causal role and intentions are relevant for assigning blame and punishment (p. 360). Additionally, in evaluating a behavior to be wrong, it is relevant whether the decision-maker himself believes in causing and striving the outcome irrespective of the actual implementation of the unpleasant outcome (Cushman 2008, p. 363). For example, an action may be classified as wrong but not blameworthy if it does not lead to negative consequences.

The Path Model of Blame by Malle et al. (2014) embeds blame judgments in a conceptual structure emphasizing the relation between blame, causality, and intention. In order to blame someone, different psychological processes take place, in which causality, intentionality, and reasoning are at the center and determine the assigned blame for an event.[4] Additionally, intentions or non-behavior, such as omissions, are also part of this approach. Two necessary conditions must be met for someone to be judged blameworthy: (1) norm violation and (2) the decision-maker's causal role. If a decision-maker is identified to be the cause of the norm violation, questions about the intentionality of his action arise, where the distinction between intentional and unintentional actions is fundamental. Furthermore, mitigating reasons of intentional acts alleviate the assigned blame. Whenever an action is classified as unintentional, it is considered what the decision-maker should (obligation) and could (capacity) have done differently. If the unintended action could have been prevented, the comparison to available alternatives determines the blame assignment (Malle et al. 2014, pp. 151-156).

Summary 2.2: Causality of the Actor

These theoretical considerations identify the causal role and the intention of the actor, his control over the action, acts of omission and commission, and norms defining the appropriate behavior to influence the assigned responsibility and may, therefore, interact with each other.

2.1.3. Causality of the Action

The second category of models focuses on formal tools such as logic, probability, and counterfactual thinking. Thus, an action's causal role, rather than the person behind the action,

[4] In their sense an event can be either an outcome (it is possible that no person is involved in an outcome) or a specific behavior (= action) of a decision-maker (Malle et al. 2014).

determines the outcome. The idea of mental simulations and the imagination of possible alternatives, proposed by Kahneman and Tversky (1982), has a lasting impact on psychological causality research. The so-called counterfactual thinking involves the (subjective) imagination of other possible alternatives the decision-maker could have chosen and the probability that the outcome would have occurred under those alternatives. A different (better) outcome that could have occurred if the decision-maker had chosen a different alternative is considered. Thus, counterfactual thinking is relevant in determining the perceived cause of an outcome (Mandel and Lehman 1996, p. 450).

When people evaluate an outcome or action based on other possible alternatives, they mentally construct a different reality, which is then used as a reference point for the actual action or outcome, even though this may lead to inappropriate evaluations (Wells and Gavanski 1989, pp. 161-167). Mandel and Lehman conclude that "in counterfactual terms, an omitted event may be mentally added and a committed event may be mentally subtracted" (p. 453). Even if the meaning between both statements is the same –"If only I had stayed awake" and "If only I had not fallen asleep" – they are described in additive or subtractive ways (Mandel and Lehman 1996, p. 453).

The following three theoretical frameworks of causality, counterfactual thinking, and responsibility attribution have been selected for their explanatory power for the subsequent analysis of this study. Each of these frameworks not only focuses on the attribution of responsibility per se but also includes grads of responsibility. This allows the earlier considerations to be extended from yes/no statements to analyses involving gradations of responsibility. Additionally, all of them focus on more than one actor, i.e., a group of decision-makers.

Spellman's (1997) crediting causality hypothesis addresses the question of how people attribute causality to multiple events that sequentially contribute to an outcome. Each sequence determining the final outcome is evaluated on the probability of changing the outcome. Then, causality is assigned based on the relative contribution of each sequence. The so-called counterfactual reasoning refers to thinking about alternatives or what might have happened if an earlier action (or sequence) had not occurred (pp. 323-327). Summarized, causality is attributed to the probability that each event could have changed the outcome. Thus, the (exact) order of an event in a chain affects the assigned causality and, therefore, responsibility for an outcome. Two different chains are distinguished: first temporal chains, in which two or more independent events contribute to the outcome. Second, there are causal chains, in which events are not independent, so that earlier events cause later ones. The attribution of causality is different between the two chains. In temporal chains,

the last event is considered more causal than the first, whereas in causal chains, the first event is seen as more causal (Spellman 1997, pp. 324-325). The idea of causal chains is not new as Kahneman and Miller (1986) have already dealt with them. However, they regard them on a more general level when it comes to multiple events that determine an outcome. The crediting causality hypothesis also explains why acts of omission are considered as less causal than acts of commission. Acting changes the probability of the outcome, so more causal responsibility for the outcome is assigned. In contrast, as an act of omission does not change the outcome probability, no (less) causal responsibility is attributed (Spellman 1997, p. 345).

However, a causal action can also be the first person in a decision chain initiating the sequence that leads to the outcome (Spellman 1997). This is quite relevant for this study, as the decision-makers decide sequentially in a chain of events. If our decision situation is interpreted as a causal chain, it would be logical to punish the first decision-maker, who chooses the unequal allocation, since he is considered more causal for the negative outcome. In this regard, the crediting causality hypothesis can explain why the first decision-maker receives more punishment points than the second and third. Additionally, if the unequal allocation is preselected and can be implemented through an act of omission, this does not increase the probability of implementing the unpleasant outcome. Consequently, less or no causal responsibility is assigned, which we specify in the following working hypothesis.

Working Hypothesis 1 (Initiation) *In a causal chain, we assume that the first decision-maker is held more responsible as he starts the decision sequence.*

Often causality is assigned in a definite way, with someone either being causal or not, so that no gradation of causation is considered. One approach that differentiates the degree of assigned responsibility is the Structural Model Approach by Chockler and Halpern (2004). Furthermore, the relationship between causality, counterfactual thinking, and responsibility is examined in more detail. In their understanding, causality depends on counterfactual thinking, which means that the outcome would not have occurred if the action had not taken place. Therefore, the authors specify that "A is a cause of B if B counterfactually depends on C under some contingency" (p. 94).

Since many outcomes depend on these contingencies, the degree of A's responsibility for B is defined as 1/(N+1), where N determines the minimum number of changes for B to be counterfactually dependent on A. More practically, if three decision-makers decide with a majority vote between an equal and unequal allocation, and they all choose the unequal one, then, the degree of responsibility of each decision-maker is 1/2, since two changes are

required to implement the equal outcome. Although not explicitly named pivotality, this definition implies that the decision-maker who finally determines the outcome is responsible for it if some other (better) action could have changed the outcome (Chockler and Halpern 2004, pp. 94-95). Since pivotality is an important concept in this study, the following definition clarifies how pivotality is defined here.

Definition 2: Pivotality

Being pivotal means determining the final outcome insofar as choosing another action would not have been implemented the outcome.

Based on pivotality as an important determinant of responsibility attribution, we formulate the following working hypothesis.

Working Hypothesis 2 (Pivotality) *We expect the pivotal decision-maker to be seen as more responsible than the non-pivotal decision-maker.*

The responsibility measure, developed by Bartling and Fischbacher (2012), predicts, that more responsibility is assigned when a decision-maker has a more considerable impact on the probability of implementing the unfair allocation. If the action does not increase the probability of an unfair outcome, it is zero, while more responsibility is assigned in proportion to an increase in this probability. Again, the decision-maker with the largest impact in determining the outcome is considered more responsible, thereby the pivotal decision-maker (p. 74).[5]

A recent work of Engl (2018) uses the Structural Model Approach as a starting point. It captures how to assign responsibility for a decision made by multiple decision-makers and focuses on an individual's causal effect in determining the outcome, the so-called causal responsibility. The individually assigned responsibility is characterized as the distance of being pivotal, i.e., the minimum number of changes (in terms of actions) to make the decision-maker pivotal (pp. 1-3). The actor's causal responsibility is determined by two forms of responsibility: ex-post and ex-ante responsibility. An agent is ex-post causally responsible if he is pivotal for the resulting outcome. The more an agent is away from being pivotal, the smaller his ex-post causal responsibility is. Ex-ante causal responsibility deals with uncertainty about the extent of ex-post causal responsibility for a given action. This means that if the actions of different actors have different probabilities in determining the same

[5]For the exact formalization see Bartling and Fischbacher (2012), pp. 84-86.

outcome, their ex-ante causal responsibility is different, even though they have the same degree of ex-post causal responsibility. Even for outcomes that are not implemented, the notion of ex-ante responsibility can assign responsibility to agents who may have been pivotal for those unimplemented outcomes, although the ex-post causal responsibility would be zero (Engl 2018, p. 3). Engl (2018) uses the experiment of Bartling et al. (2015) to underline the predictive power of his theoretical approach for sequential voting situations by emphasizing that the punishment behavior can be correctly predicted with his model (p. 28).[6]

Summary 2.3: Causality of the Action

Summarized, the previously addressed theoretical concepts focus on the action as causal responsible for the implemented outcome, where two contradictory perspectives are possible. When the decision-process is seen as a chain, the last action is considered more responsible in temporal chains while the first action is more responsible in causal chains. Alternatively, the notion of pivotality views the action finally determining the outcome as most responsible.

2.1.4. Actor and Action Causality Combined

So far, most of the theoretical research on counterfactual reasoning has focused on the resulting outcomes rather than on the criteria for assigning blame and the factors that mitigate (or reinforce) the assigned responsibility. The Structural Model Approach analyzes the relationship between counterfactual reasoning and responsibility from a theoretical perspective and describes it formally but does not take morally relevant factors into account, such as intentions or mental states of the decision-maker (Alicke 2000, p. 556; Chockler and Halpern 2004, p. 95). This section discusses two closely related theoretical frameworks that consider both the causal role of the actor and the causality of the action.

Gerstenberg et al. (2018) propose a theoretical framework that encompasses these two dimensions of causality. They consider the causal role an action played for the resulting outcome and what the action reveals about the decision-maker. Therefore, they provide an approach to make quantitative predictions about how people assign responsibility. To do this, they use Chockler and Halpern's (2004) Structural Model Approach to capture the causal role an the action plays (i.e., thus how close it is to pivotality) and extend it by a Bayesian model that allows inferences about the nature of the decision-maker. Hence, their framework builds

[6] More details on the applicability to our study is given in Chapter 7.

on two distinct processes. The first process involves an individual component: the expected behavior in the future. The decision-maker is considered less blameworthy if actions could lead to behaving correctly in the future (and vice versa). The attribution of responsibility is strongly linked to expectations, comparing actions to normative standards. The second process measures the pivotal role of the decision-maker, where the more blame is assigned the closer his action is to pivotality. Thus, the decision-maker cannot be held responsible if he is not causally related to the outcome (Gerstenberg et al. 2018, pp. 122-126).

Langenhoff et al. (2021) extend this framework by specifying the two processes in more detail. The first involves a '*dispositional inference*' about the character of the decision-maker that is drawn from his action. Based on the decision context and the decision-maker, the judging person builds expectations of how the decision-maker will act in the future. If the action and the expectation differ, other (observed) uncontrollable factors are considered so that the decision-maker is not held responsible for factors beyond his control (pp. 1-2). As before, the second process includes the causal role of the decision-maker in determining the outcome – that is the distance from being pivotal. Consequently, the more the decision-maker's action influences the outcome, the more responsibility is assigned (Langenhoff et al. 2021, pp. 1-2).

Summary 2.4: Causality of the Actor and the Action

Summarized, pivotality, the expected behavior in the future, and the decision-maker's control over his action determine the assigned responsibility.

2.2. Experimental Evidence of Responsibility Attribution in Group Decisions

2.2.1. Decision-makers' Choices

Compared to a single decision-maker, group decisions are more strategic and selfish (see e.g. Cason and Mui 1998; Luhan et al. 2009; Panchanathan et al. 2013). In a meta-study of more than 200 experimental results on dictator games, Engel (2011) shows that significantly more selfish decisions are made when the decision is shared with others (OLS regression with no treatment dummies, Coef. $-0.108, p < 0.1$) (pp. 601-602).

Experimental evidence shows that a group context increases moral transgressions because

the perceived responsibility is diffused and unethical behavior is hidden behind the group (among others, see e.g. Falk and Szech 2013; Bartling et al. 2020; Falk et al. 2020). Falk and Szech (2013) investigate that market interaction leads to decreased moral values and increased moral transgression, i.e., killing a mouse. For repeated interactions, Bartling et al. (2020) are unable to conform this result. They exhibit an erosion of moral values due to repeated interactions and not the market institution. For decisions involving moral transgression, Feess et al. (2020) find an impact of the voting threshold (the number of votes required). Sharing a decision with a group reduces one's own feeling of guilt, and most importantly, guilt is shared with others. Even if a decision-maker chooses an unfair or selfish allocation, the guilt is divided among the entire group, making it easier to act selfishly (Battigalli and Dufwenberg 2007; Rothenhäusler et al. 2013). Consequently, guilt sharing and consensual voting are identified as motives for moral transgressions.

Group composition affects behavior in that groups with more women are more kind to external parties (Cason et al. 2022) and are generally more generous and egalitarian oriented when composed of more women than men (Dufwenberg and Muren 2006). In a meta-analysis of the standard dictator game with given endowment, Bilén et al. (2021) identify women as giving on average 4 percentage points more than men, which is statistically significant.[7]

A recent experiment by Brütt et al. (2020) is related to our study in some dimensions, as it examines the diffusion of individual responsibility on the selfishness of group decisions. In a binary dictator game with groups of three, two dictators simultaneously vote on a selfish or unselfish allocation, which is implemented if both choose it. If no agreement is reached, the default allocation is implemented. The authors firstly vary the default option, either the selfish (A) or unselfish (B) allocation, so that individual pivotality cannot be diffused in the default option B (as both must vote on A to implement the selfish option). Second, the group formation process is varied in that groups are formed either exogenously or endogenously, while in the latter individuals choose if they want to decide individually or in a group with an announced selfish or unselfish default (pp. 5-6). As a result, the selfish option is not chosen significantly more often when groups are formed exogenously (A: 0.38 vs. B: 0.36, Fisher-Pitman permutation test (FFP), $p = 0.373$), whereas there is a significant difference in endogenously formed groups choosing the selfish default significantly more often (A: 0.38 vs. B: 0.25, FFP, $p < 0.01$). Thus, endogenously grouping reinforces selfish decisions when the selfish choice is presented as the default. Additionally, selfish individuals[8] choose a group

[7] In the meta-analysis of Engel (2011) women give, on average, 5.8 percentage points more.

[8] Participants are classified as selfish (respectively pro-social) depending on their choices with exogenously formed groups. Choosing the selfish option A, irrespective of the default, classifies someone as selfish while voting for the unselfish option B, irrespective of the default, means being prosocial (Brütt et al.

with a selfish default-option, where individual responsibility can be diffused, significantly more often (35%) than with an unselfish default (20%) (FFP, $p = 0.015$). In contrast, prosocial individuals are significantly more likely to join groups in which the unselfish option B is declared as default, such that the unselfish option can only be implemented if both group members choose the selfish option A (39% vs. 56%, FFP, $p = 0.002$). Consequently, the differences between endogenously formed groups are the result of a self-selection behavior of individuals (Brütt et al. 2020, pp. 10-15).

Summary 2.5: Experimental Evidence on Decision-makers' Choices

Overall, these experiments highlight that decision-makers are more selfish and are more likely to commit moral transgressions in a group setting. Pivotality is not as obvious as in an individual choice setting as responsibility is diffused. Repeated interactions further decrease moral values and selfish defaults lead to more selfish choices.

Next, we turn to the recipients side and focus on how they attribute responsibility to decision-makers.

2.2.2. Recipients' Reaction

The experiment by Bartling et al. (2015) is the closest to this study as the experimental design is adapted from it. Bartling et al. (2015) analyze individual responsibility in a sequential voting setting, in which three decision-makers decide sequentially over an equal or unequal allocation for the whole group of six.[9] The equal allocation distributes 5 points to each group member, whereas the unequal allocation awards 9 points to each decision-maker and 1 point to each recipient. The outcome, either equal or unequal, is implemented with majority vote, i.e., when at least two decision-makers choose the same allocation. After the outcome is determined, the three recipients can punish each decision-maker individually, costing them 1 point. The role assignment as decision-maker and recipient is randomly established, and the strategy method is used to assign punishment points (pp. 134-135). In our experiment we extend the design of Bartling et al. (2015) in two dimensions. First, we vary the group-building process, i.e., how participants are assigned to the roles of decision-maker and recipient, using either a legitimate or illegitimate mechanism. Second, one of the two

2020, p. 10).

[9]Bartling et al. (2015) use the expression 'voter' for the decision-maker, though the two terms are used interchangeably in this study.

allocations is preselected as a default option and is automatically implemented if no other option is actively selected.[10] The results of Bartling et al. (2015) can be summarized as follows: when the unequal allocation results, the pivotal decision-maker (the one who determines the final allocation) receives significantly more punishment points (2.12 of 7) than the non-pivotal decision-maker, i.e., the first to choose the unequal allocation (1.72 of 7), which is significantly different at the 1%-level (Wilcoxon signed rank test, $N = 72, p = 0.032$). However, estimating different OLS regressions gives a more nuanced picture. Considering the different punishment motives alone shows that being the initiator[11] influences the assigned punishment points to a higher extent (Coef. 1.604 vs. 1.494, both significant at the 1%-level) and has a stronger explanatory power (0.250 vs. 0.124) than pivotality. Including all punishment motives together specifies that the initiator and pivotal decision-maker have a significant effect on the assigned punishment points while the impact (in terms of increase in assigned punishment points) of initiation is still higher (Coef. $0.517, p < 0.05$ vs. $0.403, p < 0.05$) (p. 137). A more detailed presentation of the main results as well as a comparison with our experimental results is presented in Section 7.4.6.

Recently, Anselm et al. (2022) have extended the study of Bartling et al. (2015) in three ways. First, they increase the group size to ten (five decision-makers and five recipients). Second, they allow for reward and (or) punishment (reward/punishment only and a combination of both). Third, they use eye-tracking to examine the decision process itself. The main experimental design (voting situation, allocation, majority rule) is similar to Bartling et al. (2015), except for the group size. As in Bartling et al. (2015), the role assignment (decision-maker or recipient) is randomized and the strategy method is used to assign punishment points. During the punishment stage, process measures (response time and eye-tracking) are collected (pp. 5-7). Overall, they replicate the findings of Bartling et al. (2015), according to which pivotality strongly predicts the assigned punishment points. The pivotal decision-maker is punished significantly more than the other two intentional decision-makers (1.52 vs. 1.18 and 0.99, Wilcoxon signed-rank test, both $p < 0.02$). In contrast to Bartling et al. (2015), they identify the choice of unequal as the motive with the highest explanatory power, while pivotality and intention have the lowest explanatory power (pp. 10-14). However, they find more heterogeneity at the individual level. People fall into three categories that are stable across conditions (punishment only as well as punishment and reward): people who assign little or no punishment points (27 out of 60 participants, 45.0%), people who assign responsibility according to pivotality (11 out of 60 participants, 18.3%), and people who

[10] For a more detailed presentation of our experimental design, see Section 6.2.
[11] Bartling et al. (Bartling et al. 2015) use the term 'intention unkind'.

focus on the decision-maker's choice (19 out of 60 participants, 31.7%) (Anselm et al. 2022, pp. 16-18).

Another aspect that affects the assigned responsibility for each group member is agenda power. Duch et al. (2014) exhibit that people with agenda power (in their setting, a proposer) have more influence on the outcome and are punished more than the other decision-makers. In contrast to Bartling et al. (2015), they use a dictator game with punishment in a *simultaneous* voting setting. Depending on the treatment, group members differ in their influence on the group outcome, as they have different voting weights. These vary from equal weights for all decision-makers to unequal weights with veto power for one player. Additionally, one randomly assigned decision-maker is declared as the proposer who suggests an allocation of the endowment for the group, which is then voted on. If the proposal receives at least 51 votes, it is implemented. Otherwise, the proposer must suggest another allocation, which is again voted on. If no agreement on the proposed allocation is reached in three rounds, no one is paid, implying the decision-makers and recipients go away empty-handed. If a proposal is accepted, the recipients can punish (costless) each decision-maker individually (Duch et al. 2014, pp. 372-375). Turning to the results reveals some interesting aspects. More than three quarter of participants focus exclusively on the proposer when deciding who to punish and not on the size of the voting weight. More precisely, a proposer is punished significantly more than a non-proposer who has the same voting weight. Nevertheless, the decision-maker with the largest weight is punished more than the others, which increases in size as the size of the weight increases. Bringing both factors together, if a decision-maker is proposer and has the largest voting weight, he is almost universally punished (pp. 376-379). Thus, at the individual level, the decision-maker with proposal power is punished significantly more than the other group members. Interestingly, the decision-maker with veto power is not (significantly) punished for not using his power. Consequently, the passive action of the veto player, who does not use his veto to impose an equal allocation, is more accepted than the active proposition of an unequal distribution (Duch et al. 2014, pp. 376-381). Consequently, "individuals favor agenda power and the largest vote weight as heuristics for attributing responsibility for members of collective decision making bodies" (Duch et al. 2014, p. 388).

Another form of agenda power, as in committees, is examined in Fischbacher and Schudy's (2020) recent experiment. In a committee where members vote sequentially to approve proposals, coalitions to trade votes are often formed to make a proposal pass. Therefore, proposals early on the agenda have a higher probability of being accepted, and committee chairs prefer to get their proposals on the agenda early. In their study, Fischbacher and Schudy (2020) examine whether a committee chair can use agenda power to enforce his interests or

whether other group members punish him for exploiting his power (pp. 1813-1814). Based on the experimental voting game by Fischbacher and Schudy (2014), a committee consists of three persons who vote by majority on three independent proposals, with each committee member strictly favoring one of them. After each voting round, all players are informed whether the proposal is accepted or rejected before the next proposal is voted on. Their setting includes randomly determined sequences of proposals and sequences suggested by the chair, as well as variations in the available information about the voting behavior (pp. 1814-1816). Regardless of the treatment, the probability that the first proposal is accepted is always higher than for the second and third proposals. However, there are differences in the acceptance rate between treatments. When the chair sets the agenda, the first proposal is accepted 45% of the time with partial information and 75% of the time with full information, compared to 34% and 67% when the agenda is randomly set. Additionally, agenda control has a positive effect on reciprocity in that members are more likely to accept subsequent proposals if their own proposal has been accepted. This effect is significantly different under full information between the random and chair treatment (Wald test, $p = 0.089$). Thus, when individual voting behavior is transparent, manipulating the agenda can be profitable and is not opposed per se. However, reciprocal reactions to rejecting or accepting a proposal depend on each committee member's own benefit (or cost) of the agenda. Compared to a randomly determined order of proposals, more proposals are accepted when the chair determines the agenda using his agenda power (Fischbacher and Schudy 2020, p. 1814).

So far, these experiments have separately highlighted the exceptional role of pivotality and agenda power in assigning responsibility. Both are obvious focal points for the punishment decision and get more attention. To further investigate whether agenda power is relevant when pivotality is present, Duch and Stevenson (2014) replicate and extend the experiment of Bartling et al. (2015) by introducing a proposal stage with a randomly assigned decision-maker as a proposer. This proposer has the option to choose two out of three allocations for the following sequential voting stage. Two of the three possible allocations are the same as in the original study, while the third allocation is also unequal (7,7,7,3,3,3), but less unequal than the other allocation, which is (9,9,9,1,1,1) (pp. 10-12). Replicating the experiment yields the same results, with the pivotal decision-maker being punished significantly more than the other group members. Adding the proposal stage exhibits that agenda power is relevant for the attribution of responsibility. Especially, when the proposer chooses two unfair allocations to vote on, he is punished more than in the case of one unfair allocation (Duch and Stevenson 2014, pp. 14-24).

As the experimental findings presented here are quite important for the general punishment behavior in our study, we summarize them in Table 2.1, highlighting the key design elements and results of the experiments.

Table 2.1.: Responsibility Attribution in Groups

Experiment	Design elements	Main results
Bartling et al. (2015)	Modified dictator game with groups of six (three decision-makers and three recipients)	• The pivotal decision-maker receives on average significantly more punishment points than the non-pivotal decision-makers. • An OLS regregression reveals that being the initiator increases the assigned punishment points to a higher extent than being pivotal.
Anselm et al. (2022)	Design of Bartling et al. (2015) with groups of ten (five decision-makers and five recipients) and only punishment (respectively reward) as well as punishment and reward	• The pivotal decision-maker is punished significantly more than the other two intentional decision-makers. • Choosing unequal is the motive with the highest explanatory power for punishment.
Duch et al. (2014)	Simultaneously voting game with varying voting weights	• The proposer is punished significantly more than the other group members, as well as the decision-maker with the largest voting weight. • The decision-maker with a veto is not punished significantly more if he does not use the veto to prevent an unequal outcome.
Fischbacher and Schudy (2020)	Groups of three form a commitee and vote sequentially on the acceptance of proposals	• Having and using agenda power can lead to a favorable outcome, which is more likely to be accepted by other committee members if they benefit as well from the agenda power.
Duch and Stevenson (2014)	(1) Replication of Bartling et al. (2015), (2) a treatment where a proposer stage is added	• (1) Without proposer-stage (replication): the pivotal decision-maker is punished significantly more than the other group members. • (2) With proposer stage: agenda power affects the assigned punishment points, especially when two unfair allocations are put to a vote.

Of course, all these experiments involve a group context where decisions are made collectively by groups of people. However, the group context is still small, as group members decide anonymously and sometimes sequentially without communication with other group members. Moreover, no shared group identity is fostered, neither by declaring membership in the same group, discussed further in Chapter 4, nor by allowing communication between group members.

2.3. Political Voting and Responsibility

In the political sphere, many decisions are made jointly by groups or committees with different voting thresholds, as a result of which the evaluation of (ex-post) responsibility for implementing reforms or political decisions has long been studied. Shapley and Shubik (1954) offer a method for evaluating the spread of power in a collective political system. An individual's power is defined as the chance of being decisive for the success of a decision, thus being pivotal. A pivotal voter has the decisive vote to reach a majority and, therefore, can change the outcome of the electoral decision (pp. 787-788). The importance of pivotality in psychology and economics has already been highlighted in the previous section. In the political field, the pivotal voter is highly relevant, because he can determine the outcome of a voting process and has far-reaching implications on the political realm. If a majority has already been reached, another vote does not change the outcome and is, therefore, less important. Apart from Shapley and Shubik (1954), a large amount of research deals with the attribution of responsibility. For instance, proposal power and voting weights are discussed in a variety of ways and used to predict the influence of different policies. In particular, the influence of voting weights is a complex construct, for which various indices have been developed to measure them (e.g., Banzhaf 1964; Shapley and Shubik 1954; Penrose 1946; Gelman et al. 2002, 2004; Duch and Stevenson 2013).

In political science, economic performance is often used as a reference point to evaluate the political actions of the incumbent. There are several approaches using economic performance as a benchmark for evaluation. One approach focuses on the individual economic well-being, whereas a second approach considers the economic well-being of the (whole) nation. A third approach identifies both as relevant when voters evaluate a government, in particular the performance of a government (Kinder and Kiewiet 1981; Kramer 1983; Markus 1988). In this regard, voters use their votes to punish or reward policymakers for positive or negative economic outcomes. Voting for (against) an incumbent party is the attempt to assign responsibility to that party for having implemented or proposed good (bad) policies. Thus, the incumbent's (economic) performance is fundamental for the evaluation, rather than promises stated before the elections or while in government (Fiorina 1982; Franzese 2002; Lewis-Beck 1988, Chapter 6). In addition, political institutions are also crucial for the attribution of responsibility. The stronger a governing party is, the more the economic circumstances matter for the support of that party or government (Anderson 1995; Powell and Whitten 1993).

If a government consists of several parties, it is less likely to be punished (for misgovern-

ment) than individual parties. Because responsibility for (economic) outcomes is dispersed, the incumbent and the opposition may obscure their roles, making voters' assessments less clear (Powell and Whitten 1993, pp. 392-393). Under the assumption of complete information about responsibility attribution across parties, Duch and Stevenson (2008) find out that the more the responsibility is concentrated on one party, the higher the level of economic voting is. Furthermore, the possibility of shared responsibility in a multi-party system plays an important role in determining economic voting behavior.

Being the incumbent party or government is beneficial in that it increases the probability of re-election (among others, see e.g. Quattrone and Tversky 1988; Carey et al. 2000; Aytaç 2018). Even when the economic policy of two political parties has the same expected value, so that voters are indifferent between them, they prefer voting for, or rather stay with, the incumbent (Quattrone and Tversky 1988, pp. 724-725). Additionally, an incumbent benefits from good economic outcomes, while poor economic performance can negatively affect reelection. More specifically, if economic results are good, the incumbent stands for this positive development and is viewed as continuing the positive trend. In contrast, the election of an (unknown) challenger is associated with more risks. However, it may be chosen if the economy has suffered losses under the incumbent, although it is entirely unclear whether the challenger achieves better results (Quattrone and Tversky 1988, pp. 724-726).

There is a wide range of empirical research in the political domain on the measure of the advantage of being the incumbent in state legislative elections. For example, when data from the US elections between 1968 and 1986 is used, incumbency is found to have an increasing vote-related advantage as legislators have larger budgets to facilitate their service activities. Consequently, incumbency has an advantage in the form of higher vote shares (among others, see e.g. King 1991; Cox and Morgenstern 1993). Carey et al. (2000) develop a model estimating the value of incumbency in terms of the most relevant variable – the reelection probability. Using data from the US-legislative elections between 1992 and 1994, they identify a high value of incumbency in all 96 chambers, with 68 of 96 chambers having an incumbent's probability of winning greater than 90% (pp. 677-682).

Some (empirical) studies examine the effect of relative international performance on the incumbents' advantage in reelecting (among others, see e.g. Kayser and Peress 2012; Campello and Zucco 2016), while Aytac (2018) also captures the domestic effect of economic performance. Analyzing data from 475 national-level elections in 62 countries between 1965 and 2014, he hypothesizes that voters compare their own country's economic performance with the performance of other countries. If their own country performs better (worse) the government is (not) reelected (pp. 17-22). Using OLS regressions the positive effect of relative

domestic and international growth on voting for the incumbent is confirmed. More specifically, a one-percentage point increase in relative domestic growth increases the incumbent's vote share by about 0.57 percentage points ($p < 0.01$), while relative international growth increases the vote share by 0.72 percentage points ($p < 0.05$) (Aytaç 2018, pp. 23-24). Evaluating economic performance usually involves accessing such information (e.g., through media consumption) as well as handling and understanding it. This access and consumption of current news is positively correlated with voters' prior knowledge (Price and Zaller 1993) and their level of education (Jerit et al. 2006). To control for this, the interaction of the average year of schooling with economic performance is included in the regression analysis. A positive and statistically significant effect ($p < 0.1$) is found for relative international growth, implying that education increases its influence on the choice of incumbency once a threshold of at least eight years of schooling is exceeded (Aytaç 2018, pp. 24-25).

In this respect, Gomez and Wilson (2003) analyze the attribution of responsibility to the Congress and the President in the United States of America by focusing on the importance of the sophisticated voter. When political sophistication is low, the most apparent political figure, like the President, is held responsible for political decisions. In contrast, people with higher levels of sophisticated thinking are able to understand the complex political processes that lead to policy outcomes and distribute responsibility among multiple political actors. Furthermore, members of the incumbent party, rather than all congressional members as such, are held responsible for the economic performance (pp. 276-281). Duch and Stevenson (2008) point out that voters take minor information and observable conditions into account when they evaluate the performance of a government, which are then used as cues for potential future governments.[12]

The value of incumbency is explainable by loss aversion since a potentially negative outcome (here, the risk of voting for the challenger and no improvement in economic performance) is seen as worse than the positive outcome if the challenger wins and the economic performance improves. A preference for the status quo over other alternatives, which have the same expected value, is a direct consequence of loss aversion. Therefore, the incumbent faces another advantage: he is in the status quo position. His status quo policy is the reference point against which the policy of another party (the challenger) is compared. People generally tend to stick with the status quo, so the incumbent is in a better position to maintain his mandate (Quattrone and Tversky 1988, pp. 724-725). This bias toward the status quo can negatively affect the efficiency of reforms if winners and losers cannot be identified

[12]Healy and Malhotra (2013) provide an overview of recent studies of how the governments' performance is evaluated by voters.

before implementing a reform (Fernandez and Rodrik 1991). Section 3.1 further addresses the status quo bias and loss aversion in light of this study.

Summary 2.6: Political Voting and Responsibility

From a political perspective the pivotal voter is identified as a focal point for the attribution of responsibility. Additionally, the incumbent has an advantage over the opponent in the form of a higher chance of re-election and a larger share of votes, especially after a good (domestic) economic performance.

3. The Economic Perspective on Acts of Omission, Social Norms and Preferences

In this chapter, we turn to an economic perspective on the decision-making context and address four relevant aspects: (1) acts of omission and the default effect, (2) social norms, (3) preferences, and (4) procedural justice. They are not new here, as the theoretical considerations about the causality of the actor has already specified them as influencing the assigned responsibility.

Section 3.1 starts with presenting theoretical and experimental evidence on acts of omission and the default effect. The relevance of this results from the fact that the causal relationship and the underlying intentions through an act of omission are less evident than through commission (Shultz et al. 1981, p. 241). Additionally, the default, an already preselected option, increases the probability that this alternative is chosen (for a meta-analysis see Jachimowicz et al. 2019). Thus, acts of omission and defaults provide an interesting opportunity to investigate the impact of causality as well as the situational context itself. The first section closes with possible explanations and reasons why people do not (always) behave as predicted by rational choice theory.

Since there is no human society without social norms, almost all aspects of human life are influenced by norms that specify how people should behave (Bernhard et al. 2006, p. 217). In this respect, social norms prescribe the appropriate behavior depending on the situational context. Norms are social as they are part of the interaction with other people or groups, who share and follow the same social norms (Fershtman et al. 2012, p. 140). In all societies, normative standards of behavior are enforced through formal and informal sanctions, while the importance of these sanctions for (economic) interaction has been shown extensively (among others, see e.g. Ostrom 2000; Fehr and Gächter 2000a; Carpenter and Matthews 2009). As we expect social norms to determine the punishment decision, various dimensions and implications are presented in Section 3.2.

The third section focuses on preferences and distinguish three types of them: (1) social preferences, (2) risk preferences, and (3) distributional preferences. Many experimental results, such as giving behavior in the dictator game, rejection rates in the ultimatum game,

and costly punishment, cannot be explained by selfish utility-maximization (among others, see e.g. Fehr and Gintis 2007; Akbaş et al. 2019). Thus, we present several models of social preferences which aims to explain such behavior. Afterwards, we turn to risk preferences and present evidence on people's risk attitudes in order to conduct expectations for (punishment) behavior in our study. Fairness is a relational and heterogeneous construct that requires comparisons with other individuals, groups, or societies (Andreoni and Miller 2002). Although not everyone has the same perception and value of fairness, it is essential to many people. They are willing to give up money in order to be consistent with or avoid substantial deviations from their justice principles supporting their far-reaching implications (Cappelen et al. 2007, p. 818, 2010, p. 441). As distributional preferences and justice considerations may influence the assigned responsibility in this study, a brief overview of them is provided in Section 3.3.3.

The last Section 3.4 concentrates on procedural fairness. Procedural fairness specifies that a decision is evaluated not only on the basis of the outcome, but also on the fairness of the procedure that brings about the outcome (Bolton et al. 2005, p. 1054). The introduction of procedural fairness has strengthened the idea that equality means not only equality of outcomes, but also equality of opportunity, i.e., of the procedures that lead to those outcomes. The mechanisms or procedures typically used to assign participants to different roles or positions strongly influence how people evaluate the situation and how they decide (among others, see e.g., Bolton et al. 2005; Ku and Salmon 2013; Akbaş et al. 2019). Therefore, procedural justice is the segue to Chapter 4, where we address status differences.

3.1. Defaults and Acts of Omission

Technically, the difference between omission and commission is irrelevant as the outcome is the same. The Consequentialist Rational Choice Theory (RCT) does not assume a difference in behavior between actively and passively implemented outcomes as the end state is similar (Vaal 1996). However, the causal structure of omissions is often more complex or obscure because it may involve other circumstances. In addition, the intention of the decision-maker is less evident because no actual choice is made in doing nothing (DeScioli et al. 2011a; Malle et al. 2014).

As the distinction between status quo, omission, and default is not always apparent, definitions of these concepts are provided in Section 3.1.1. Additionally, Section 3.1.2 discusses experimental results and consequences of the omission bias and default effect. This section closes with an overview of explanations for the results presented (Section 3.1.3).

3.1.1. Definitions

An *act of omission* is defined as inaction, that is, an (expected) action that is not performed by a decision-maker (Shultz et al. 1981, p. 241).

Definition 3: Act of Omission

An act of omission is understood to mean that a decision is implemented passively by doing nothing.

When people have to decide something unpleasant, they often prefer acts of omission to acts of commission, i.e., they do nothing instead of actively deciding. The so-called omission bias refers to a general preference for omission over commission, which has been documented in a vast range of research, especially for unpleasant decisions (Spranca et al. 1991, pp. 76-77). In this sense, acts of omission are preferred to avoid responsibility for harmful outcomes, as someone is often only held responsible (either by himself, others or both) for acts of commission. With an act of commission, the decision-maker's intention becomes clear, which is less evident with acts of omissions. Even though omission and commission are treated differently by some people, many do not make these differences (Spranca et al. 1991, p. 77).

A *default* usually represents the existing state or current situation. Often, changing the default involves a trade-off and is associated with more uncertainty (Johnson and Goldstein 2003, p. 1338).

Definition 4: Default

The default is the preselected option, which is implemented if no other option is selected.

From the perspective of a rational individual, the default should make no difference because a rational agent always chooses the best option maximizing his payoff. If the default does not represent his (real) preference, he should choose another option. However, reality has shown that people are more willing to accept an option if it is offered as a default, which is known as the default effect (Johnson and Goldstein 2003, p. 1338; Thaler and Sunstein 2003, p. 176). Consequently, a default increases the probability that the so-presented option is selected. As defaults are easy to implement, they are used in many decision contexts to influence individuals' behavior. In a meta-analysis of the default effect on environmental, consumer and health decisions, Jachimowicz (2019) points out that the default option is

chosen 27.24% more often than the non-default option (p. 174).

Often, doing nothing implements or preserves the *status quo*, so classifying an option as status quo has an impact on whether that option is implemented. Even with new alternatives, people tend to preserve the status quo, which is called the status quo bias (among others, see e.g. Samuelson and Zeckhauser 1988, pp. 7-8; Johnson et al. 1993; Gärtner 2018).[13]

Definition 5: Status Quo

The current state of a situation or decision is the status quo, which was chosen or implemented in the past.

As a clear distinction between default, status quo, and omission is fundamental to this study, Figure 3.1 visualizes and delineates the three constructs.

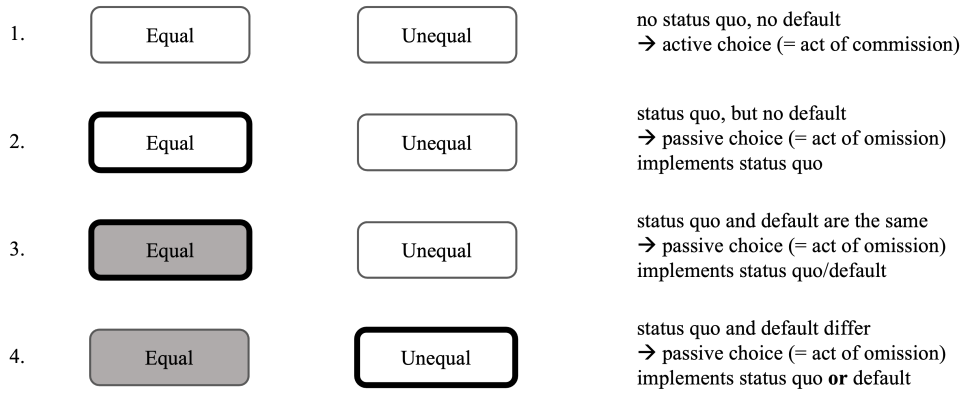


Figure 3.1.: Distinction Between the Three Different Constructs

As in this study, an individual has to choose between two possible allocations, an equal or an unequal allocation. Each situation varies the existence of a default and status quo option. In the first situation, no option is declared as status quo and preselected, so that the decision-maker must actively decide which option to implement. In the second situation, the

[13]The seminal paper of Samuelson and Zeckhauser (1988) examines the status quo bias in various (hypothetical) questionnaires and field studies. The questionnaires illustrate different decision problems ranging from funds allocations to improving of auto insurances. An option is chosen significantly more often when it is declared as the status quo than when it is framed neutrally. The two field studies examine the role of the status quo in two real world periodic decision environments: retirement and health plan decisions. Consistent with the results of the questionnaires, these studies highlight the existence of status quo bias. People tend to retain with the established option even though switching was free and easy. In summary, an alternative presented as the status quo has a higher probability of being chosen than a neutral framing. The more options are available, the more likely the status quo is chosen (Samuelson and Zeckhauser 1988).

equal option is declared as status quo (e.g., if it was chosen in the previous round), which is passively implemented by an act of omission if no active choice is made. In this situation, no option is preselected as default. Third, the decision context is extended by a default option as the equal allocation is preselected. The fourth and final situation involves a status quo and default option, but each is associated with a different allocation. Equal is declared as status quo, while unequal is preselected. If the decision-maker does nothing, the default option is implemented.

Summary 3.1: Defaults and Acts of Omission

The essential characteristic of a status quo is that it has already been implemented in the past, while a default is a way alternatives are presented or framed. The default can be considered the suggested option, which is not necessarily the status quo. Furthermore, both can be implemented either actively (act of commission) or passively (act of omission). Usually, doing nothing (which is equated to an act of omission) maintains the current state and (or) implements the default. To change the status quo or default, an active decision (i.e., act of commission) is required, whereas the omission bias is characterized by inaction (Ritov and Baron 1992, p. 50; Gärtner and Sandberg 2017, p. 2).

Generally, as many studies confound omission bias and status quo bias and use them in the same way, a clear distinction between them is challenging because an act of omission often preserves the status quo. Using hypothetical scenarios where outcomes are the same but result either from action or inaction, Ritov and Baron (1992) attempt to disentangle the intertwining of omission bias and status quo bias. More specifically, they examine whether one bias causes or supports the other or whether one bias has a more substantial influence on the other. They highlight that the omission bias exists regardless of a change in the status quo through omission or commission. This supports the assumption that the omission bias is essential for determining the status quo bias. However, unlike Ritov and Baron (1992), Schweitzer (1994) cannot support that one bias is superior to the other, although both appear to be closely related and often occur together.

In this study, the expression default is used because one option is preselected and implemented by an act of omission. When time runs out and participants do not actively confirm or change their choice, the preselected option is realized. Consequently, it is more appropriate to use the term default than status quo. Furthermore, as each round begins

independently of the previous round, no current state (= status quo) is apparent and the allocation implemented in the last round is not declared status quo.

3.1.2. Experimental Evidence

Several studies examine the omission bias and focus on how omissions and commissions are judged as well as factors explaining the bias. The intention and the causal role of the decision-maker are highlighted as one reason why acts of omission are judged less harshly than acts of commission (Spranca et al. 1991; DeScioli et al. 2011a; Malle et al. 2014).

The pertinent vignette study by Spranca et al. (1991) uses six hypothetical scenarios to demonstrate the existence of the omission bias. Additionally, they focus on why it is judged differently examining potential reasons and explanations – for example, the first case in the first experiment, in which students at the University of Pennsylvania must judge the morality of John. John harms his competitor Ivan either by failing to inform Ivan about an allergic ingredient in the dish he orders, or by actively recommending the allergic dish before (after) Ivan makes his choice. Regardless of John's action, the unpleasant outcome is the same: Ivan eats the allergic dish and becomes ill. 65% of subjects judge the harm caused by omission less bad than the same harm caused by commission, which is significantly different at the 5%-level ($\chi^2 = 4.78, p < 0.025$, one-tailed) (pp. 82-84). Summarizing the results of the different scenarios/experiments leads to the conclusion that participants strongly prefer to implement harmful outcomes through omission. They are less blameworthy (by themselves and others) as the intention and knowledge of the resulting consequences and their causal role are less pronounced in omissions (pp. 81-103). Kordes-de Vaal (1996) supports these findings in a similar design concluding that the perception of the agent's causal role affects the judgment and assigned responsibility for the outcome. It appears that the basis of the omission bias is a difference in the actor's perceived causality. The causal link between an act of omission and its consequences is less intense, which makes an omission's outcome appear less intentional than an commission's outcome (pp. 169-170).

Further experimental evidence is given by DeScioli et al. (2011a) and DeScioli et al. (2011b), who focus on the judgment of omissions by third-party observers. Using hypothetical scenarios, where the behavior of a person relates to the death of another person, DeScioli et al. (2011a) carry out four different experiments where undergraduates have to rate the moral wrongness and deserved punishment of the person. In their experiments they use the same two hypothetical scenarios but vary the degree of causality and transparency holding intentions constant. A '*do-nothing*' button without causal impact on the outcome is seen as

more wrong than actually doing nothing, even if the outcome is the same (pp. 206-213). Relating to this, DeScioli et al. (2011b) use a take-game on Amazon Mechanical Turk in which a first mover can take money from a second mover either by actively choosing to do so or by refraining. Depending on the treatment, an unaffected third party can punish the first mover by reducing his payoff. Although omissions are less efficient, a strategical use of omissions is identified when punishment is possible, while they are also less likely to be punished than acts of commission. More precisely, in the no-punishment condition 28% take the money by an act of omission while 51% do so in the punishment condition ($z = 2.24, p < 0.05$), which is also significantly less punished ($t(40) = 2.93, p < 0.01$) (DeScioli et al. 2011b, pp. 442-445).

Using a binary dictator game, Gärtner (2018) investigates the prosociality of intuitive decisions under time pressure where a status quo is either prosocial, pro-self, or absent. When the status quo is prosocial, prosocial choices are made faster, whereas selfish choices are made faster when the status quo is selfish. Thus, the status quo fosters a fast choice in accordance with the status quo option, highlighting the influence of a status quo on intuitive decisions (pp. 127-129). Although this design examines the effect of a status quo, the same results could be obtained with a default option. This would imply that decisions following the default are made faster, so the timeout option and automatic default implementation may not be relevant. Interestingly, this is the case in this experiment, which is discussed in more detail in Part II of this study.

Another laboratory experiment worth noting highlights that the default does not imply that people always maintain or choose the default. On the contrary, the default can help to reduce the uncertainty of a decision and consolidate the decision-maker's preferences by, for example, evaluating the plausibility of a default. Thus, presenting and labeling options as default, (may) influence the decision-maker's preferences (Dhingra et al. 2012, pp. 69-70). In their dictator game, participants play over four rounds choosing an allocation of 10 points. All possible splits of integer amounts are presented in a randomized order with the first option as default. The preselected option is one of three possible allocations (10-0, 0-10, 5-5) and changes after each round (Dhingra et al. 2012, pp. 72-73). The authors emphasize that the default in the first round affects the generosity in all subsequent rounds. If the default in the first round is fair or highly fair, i.e., 5-5 or 0-10, giving behavior in subsequent rounds is significantly higher. More specifically, participants give on average \$2.71, respectively \$3.14 with a fair or very fair default, while the average amount given in the selfish default condition is \$1.47 and \$2.09 without a default, which is significantly different ($t(78) = -2.74, p = 0.008$). However, this does not imply that the preselected

default is (always) chosen. In fact, the default can be seen as a starting point or anchor that conditions the decision-maker's preferences (Dhingra et al. 2012, pp. 73-75).

Apart from laboratory experiments, there are also many other (real-world) circumstances in which the default effect and difference between omission and commission are studied – for example, in a questionnaire study by Johnson et al. (1993), which examines (among other things) the status quo and the default effect of auto insurance. Either a full-priced policy with a right to sue for any injuries or a lower-priced insurance with sue restrictions is presented as the status quo. Participants evaluate the right to sue as more important and valuable when it is already in place (53% versus 23% when it is not implemented, $p < 0.001$), while neutral framing is intermediate in that 48% prefer the right to sue. Consequently, the right to sue depends significantly on how the option is presented, either as already implemented, neutral, or as an active choice. A $\chi^2 - test$ ($p < 0.006$) and a Kruskall-Wallis rank-order test ($p < 0.0001$) reveals a significant different distribution (pp. 46-48). A real-world insurance switch in two US states supports these hypothetical findings as changing the default option happens quite rare in both states (Johnson et al. 1993, p. 48). This example shows, that using a default or status quo can have implications and consequences in everyday life, as people often adhere to or choose the default option. In the political context, a default may be seen as a recommendation by the policymaker and thus the appropriate action to choose (McKenzie et al. 2006, p. 414). Alternatively, the default can be seen as a legitimate option, based on deliberate consideration by experts, and therefore the best option for most people (Thaler and Sunstein 2003, p. 177). Besides the political implications, there is a large body of research on the role of defaults in health situations, particularly insurance decisions (see e.g., Johnson et al. 1993), pension savings (see e.g., McKenzie et al. 2006; Carroll et al. 2009), organ donations. (see e.g., Johnson and Goldstein 2003; Gimbel et al. 2003; McKenzie et al. 2006) or environmental situations (see e.g., Pichert and Katsikopoulos 2008; Sunstein and Reisch 2021).

Summary 3.2: Experimental Evidence on Omission versus Commission

These experimental results highlight that clear evidence of wrongdoing for omissions is more difficult to provide as it is characterized by the absence of an action. Moreover, there is less evidence of the decision-maker's underlying intentions, which, as argued earlier, are essential for assigning responsibility. As a result, people strategically prefer acts of omission over commission to avoid potential sanctions in the form of punishment.

Based on these experimental results, we can formulate our next working hypothesis.

Working Hypothesis 3 (Choosing the Default) *We assume choosing the default, as it can be done by an act of omission, to be punished less than choosing the option that is not preselected. This effect could be even stronger if an unpleasant outcome is declared as default and passively implemented.*

Ignorance

As an act of omission can also result from (willful) ignorance. Experimental evidence underlines the idea that ignorance reduces the decision-maker's responsibility if he decides to stay ignorant about the actual consequences insofar as he is punished less (among others, see e.g. Conrads and Irlenbusch 2013; Bartling et al. 2014a; Wieland 2016).

The experiment of Dana et al. (2007) highlights that many people avoid informing themselves about the actual consequences of their decisions (although the information might be free) in order to act self-interested and use the so-called 'moral-wiggle-room'. Exploiting the 'moral wiggle room' makes the relationship between decisions and their consequences more ambiguous, while this lack of knowledge about the impact on the outcome reinforces selfish choices. Compared to a baseline treatment, where the exact payoffs are known, decision-makers choose the selfish option in 63% of the time (26% in the baseline treatment), with the proportions being statistically significant different at the 5%-level ($\chi^2(1) = 4.64, p < 0.03$). Even when uncertainty can be resolved, most decision-makers (56%) stay ignorant (pp. 74-75). Thus, people face a trade-off between their own utility and the welfare of others. Staying ignorant can resolve this trade-off or give them a way of justifying their self-interested behavior (Bartling et al. 2014a). Replicating and extending this work, Grossman (2014) varies the default information to investigate whether participants exploit the 'moral wiggle room' when they must choose actively to be or remain ignorant. Nearly half of participants use the original 'moral wiggle room' (clicking to become informed), a quarter stays ignorant, whereas only 3% decide actively to remain ignorant. Consequently, the default affects the decision to stay ignorant, which almost disappears when passive implementation is impossible (pp. 2660-2663). The default is seen as the experimenter's suggestion and thus legitimizes (selfish) behavior, so that ignorance reduces one's own responsibility and contributes to the maintenance of a good self-image (Grossman 2014, pp. 2663-2664).

Wieland (2016) investigates the role of moral blameworthiness for (strategic) ignorance and highlights (theoretically) that agents are blameworthy to some extent when they decide to stay ignorant, but less than fully informed agents. In a modified dictator game, where

a random draw determines the initial state with equal probability, Bartling et al. (2014a) address whether ignorance can avoid punishment for an unpleasant outcome. Regardless of the actual state, one option (a_1) always attributes (70) and the other option (a_2) attributes (50) to the decision-maker, while the payoff for the recipient is opposed and either high (50) or low (10) depending on the state and chosen option (pp. 514-516). Focusing on costly third-party punishment (i.e., punishment by an unaffected observer), they find that ignorance can prevent punishment when the resulting outcome is unfair. If a decision-maker chooses to remain ignorant, so that he does not know which state he is in, and chooses option (a_1), i.e., the higher payoff for himself he is punished significantly less when the recipient gets the lower payoff (10) compared to a situation when payoffs are known, as in the revealing situation or baseline treatment (11.42 vs. 16.25, respectively 19.72). A wilcoxon sign-rank test reveals that this difference is significantly different at the 1%-level (respectively 5%-level). However, the decision to remain ignorant is punished regardless of the fairness of the outcome, as in a mutually favorable outcome where ignorant dictators are punished significantly more than revealing dictators. (8.00 vs. 2.76, wilcoxon sign-rank test, $p = 0.034$) (Bartling et al. 2014a, pp. 516-518).

Summary 3.3: Ignorance

So far, these results have been consistent with the general preference for accepting the status quo, choosing the default, and acts of omission. If the decision-maker does not recognize that he has to make a decision or if he has a preference for the status quo, this inaction can be perceived as ignorance or decision avoidance, even though it is not intended as avoidance. Again, the underlying intention of the decision-maker is less clear, so that the attribution of responsibility is not as obvious as with active decisions revealing the decision-maker's true intention. The possibility of staying ignorant raises the question of whether the decision-maker is responsible or blameworthy for an unpleasant outcome. When the consequences of a decision are less obvious or the decision-maker is unaware of them, the responsibility for the decision is imprecise.

Delegation

Decision avoidance or delegation is a second interesting component of non-decision-making, where individuals try to avoid the responsibility for making an active decision, either by postponing, inaction, or delegating (Anderson 2003, pp. 139-140). In the political field, delegating a decision to shift the responsibility for the consequences and avoid an unpleasant

decision to an agency or a committee is quite popular (among others, see e.g. Fiorina 1982; Vaubel 1986).

Apart from the political sphere, delegation is also widespread in companies, where consultants are entrusted with the execution of unpleasant decisions. In doing so, blame and responsibility are shifted to someone else (Bartling and Fischbacher 2012, p. 68). From a rational perspective, a decision is delegated to increase overall efficiency, as an agent is more specialized and requires less time or effort (Hamman et al. 2010, p. 1826). Recently, however, several researchers have experimentally tested the influence of delegated decisions on the attribution of responsibility.[14] They highlight that delegating an unpleasant decision reduces the assigned responsibility in the form of less punishment points. In addition, delegation decreases the perceived guilt for an unpleasant outcome, even if the delegate can only choose an unfair allocation (Bartling and Fischbacher 2012; Oexl and Grossman 2012). With an existing delegate, rejection rates in an ultimatum game are lower for unfair offers (Fershtman and Gneezy 2001), while lower offers appear more often in a dictator game (Hamman et al. 2010).

Bartling and Fischbacher (2012) investigate how responsibility for delegated decisions is attributed. A decision-maker chooses between an equal allocation (5 points to each) or an unequal allocation (9 points to decision-maker and delegate and 1 point to recipient). Alternatively, he can delegate the decision to a delegate, who must choose one of the two allocations. At a cost of one point, the recipient can assign punishment points to the decision-maker and (or) the delegate, which is used as a measure of responsibility. As expected, delegating the decision leads to significantly lower punishment for the decision-maker and higher punishment for the delegate. If the decision-maker chooses the unfair allocation himself, he receives on average 4.27 points. In contrast, if he delegates the decision and the delegate chooses unfair, the decision-maker receives 1.31 points, while the delegate receives 3.96 points. Varying the decision situation (delegation to a die or playing repeatedly) confirms these results (pp. 67-76). Oexl and Grossman (2012) extend this experiment by eliminating the fair allocation for a delegated decision, allowing the decision-maker to choose the unfair allocation indirect through the delegate. Even in this setting, the decision-maker is punished less than the delegate (although delegating implies an unfriendly intention), who cannot influence the outcome.

[14]For a review of the principal-agent literature see, for example, Bolton and Dewatripont (2005).

Summary 3.4: Delegation

Summarized, these experiments highlight that delegation can shift responsibility from a first mover to a delegee. Delegating the decision right to transfer blame is used to avoid punishment.

Although the decision-maker is not the focus of this study, his decision behavior can provide meaningful insights into the role of punishment and responsibility attribution. Particularly, the second decision-maker, who opts for equal when the first decision-maker chooses unequal, may be punished more for his equal choice, because he avoids the final decision.

3.1.3. Possible Explanation

There are many potential reasons and constructs explaining the experimental results presented in this section. Apart from rational reasons, people can have an 'information leakage' (so there might be normative reasons for choosing the default), or the transaction costs to choose another option are too high (Dhingra et al. 2012). If people are loss averse and see the default as the reference point, giving up the default can be experienced as a loss (e.g., Kahneman and Tversky 1982; Spranca et al. 1991). People take possible alternatives into account (e.g., Kahneman and Miller 1986; Ritov and Baron 1992), try to avoid regret (e.g., Bell 1982; Loomes and Sugden 1982), or may just be inattentive and forget to change the default option (e.g., Thaler and Sunstein 2003; McKenzie et al. 2006). Thus, loss aversion, norm theory, and regret avoidance are prominent explanations for the emergence of an omission bias that need not be mutually exclusive. Additionally, the default effect may occur because of people's inertia, or it may be viewed as the suggested or recommended option. In what follows, we discuss these possible explanations in the light of our experiment and provide an overview of why people stick with the default, choose the status quo, and prefer acts of omission over commission.

Rationality

Maintaining the status quo, choosing the default, or doing nothing is rational if, for example, the preselected option is preferred or switching is (too) expensive. Whenever the costs (either as monetary or as transition costs) are higher than the expected gain from switching, it is perfectly rational to maintain the status quo or choose the default. Additionally, switching behavior may involve uncertainty or risk about the resulting consequences. However, as there

are experimental studies where neither transition costs nor uncertainty are essential, (i.e., in the hypothetical situations in Samuelson and Zeckhauser 1988), rationality cannot fully explain experimental findings and the general preference for the status quo or default option (Samuelson and Zeckhauser 1988, pp. 33-35; Schweitzer 1994).

Loss Aversion

The prominent statement "losses loom larger than gains" (Kahneman and Tversky 1979, p. 279) describes the main implication of loss aversion very well. Many people are loss averse and evaluate a loss more highly than a gain of the same amount. Moreover, gains and losses are evaluated depending on a reference point, which is an already adapted state, situation, or the endowment, but can also be an imagined option (Kahneman and Tversky 1982, pp. 160-162).

If the current situation is considered as the reference point, changes are evaluated with respect to this reference point, from which a relative loss is perceived as more substantial than an equivalent gain. Because people value losses more than equivalent gains, this loss-averse attitude can shape the behavior toward the status quo or default option (Kahneman et al. 1991, pp. 197-198; Ritov and Baron 1992, p. 60). Depending on the situational context, acts of omission may be viewed as a point of comparison or, rather, as a point of reference. If an act of omission is perceived as the reference point, the worse outcome is evaluated as a foregone gain. An act of commission that leads to the same worse outcome is valued as a mere loss and weighted more heavily than the same outcome by omission. Even if a good outcome results, it is evaluated more favorably by omission, as it is perceived as a foregone loss, whereas the same outcome through commission is valued as a mere gain. Thus, according to loss aversion, outcomes achieved by acts of omission are always considered better than the same outcomes by commission (Spranca et al. 1991, pp. 77-80).

Gächter et al. (2021) compare loss aversion for riskless endowment effect experiments with risky choice tasks and exhibit that loss aversion does exist in both situations (82% vs. 71%) and is positively correlated. However, published studies present mixed results concerning the occurrence of loss-aversion (among others, see e.g. Schmidt and Traub 2002; Erev et al. 2008; Trueblood 2015). Testing the extent of loss aversion in an experimental lottery setting, Schmidt and Traub (2002) classify 33% of their sample as strict loss averse, 24% as loss seeking and the rest as unclassified (within-subject analysis using a binomial test, $n = 45$).[15] The between-subject analysis for all choice situations reveals that only half of choices align

[15] A t-test where 5 observations are excluded as they behave inconsistent, leads to similar results.

with strict loss aversion, so the null hypothesis of loss neutrality and loss seeking cannot be rejected. Consequently, there are some people who can be classified as loss averse. However, there is nearly the same amount of people who are loss seeking while many people cannot be classified in either category, as a result of which loss aversion cannot be considered as generally existent (pp. 240-244). Trueblood (2015) investigates reference-dependent effects in perceptual tasks (decisions about the size of rectangles) and compares them to a consumer choice task (choosing cell phones). While the latter does include a gain and loss frame it is absent in the perceptual task. She finds out that all effects exist also in the perceptual task and that their occurrence cannot be explained by loss aversion alone. The contextual situation is used as a different explanation insofar as the reference point can function as a stimulus making similar options more likely to be chosen (p. 21).

Applying loss aversion as a potential explanation for the behavior in this study means that an unequal default may serve as the reference point for the decision-makers and recipients. Then, choosing the equal option is perceived as a potential loss, as the probability of receiving less is higher. However, it is not clear whether participants perceive the default (respectively the 9 or 5 points) as part of their endowment and, thus, as a reference point. Additionally, it is also possible that recipients use the implemented outcome of the previous round as a reference point – for example, if the equal outcome was implemented in the previous round, while the unequal allocation results in the current round.

Norm Theory

The main idea of Kahneman and Miller's (1986) norm theory is that norms are not formed in advance but after a specific event. Then, each event is evaluated based on other available alternatives that can lead to a better outcome. As the alternative to an active decision is easy to imagine, it leads to a stronger (emotional) reaction than a passive decision, where (counterfactual) thinking of possible alternatives is more complex (p. 136). The previously presented experimental results showing an omission bias are consistent with norm theory. The evaluation of an active decision is made conditional on the resulting outcome. A good or pleasant outcome leads to a more positive evaluation of the active decision, while a harmful or unpleasant outcome is seen as more negative than the supposed outcome by omission. In contrast, an act of omission is judged from a neutral perspective that does not depend on the (positive or negative) outcome (Ritov and Baron 1992, pp. 50-60).

Regret Theory

Regret can be seen as an additional element of norm theory, as "norm theory predicts omission bias because individuals anticipate more potential regret as a result of commissions, and they incorporate regret avoidance into their valuation of the options" (Anderson 2003, p. 143). In situations involving uncertainty, people may find out that another alternative (which they do not choose) has involved a higher payoff or utility for them and as a result of which negative feelings may occur. Thus, people sometimes regret results or consequences of past decisions and they try to avoid such unpleasant feelings in the future (Bell 1982). In a situation where a person has to decide between two actions A_1 and A_2, each associated with a different outcome O_1 and O_2, comparing the received outcome with the possible other outcome determines the perceived pleasure with the received outcome. For example, if action A_1 is chosen so that O_1 results, O_1 is compared to the other outcome O_2. If the received outcome O_1 is better than the other outcome O_2, so that the received result is more desirable, a positive feeling of pleasure may occur. Reversely, a feeling of regret occurs if the received outcome O_1 is worse than the other outcome O_2 that would have occurred with action A_2 (Loomes and Sugden 1982, p. 808).

Consequently, changing the status quo or default may be more remembered than maintaining it. One's perceived responsibility for the outcome may be more extensive when changing the default than when maintaining it. Outcomes that cannot be undone lead to more regret, especially if they are the result of an active decision. If a decision causes more regret because one actively chose an outcome than if one passively implemented it, individuals may try to avoid regret in the future (DeScioli et al. 2011b, p. 442). In this regard, regret is determined by counterfactual thinking, in which a decision-maker believes that he could have achieved a better outcome with a different choice. As a result, negative emotions and the desire to avoid such regrettable consequences in the next (similar) situation arises. Even if the decision is judged to be correct given the available information, maintaining the status quo can be seen as a way to avoid possible regret (Samuelson and Zeckhauser 1988, p. 38; Anderson 2003, pp. 143-144).

A large body of research suggests that regret is lower for preserving the status quo than for changing it. Thus, choosing the status quo or doing nothing can reduce feelings of regret when an unpleasant outcome occurs (Kahneman and Tversky 1982, p. 160). Although most studies focus on the status quo rather than the default, the same psychological issues apply to a default. Choosing the default may lead to less regret in the event of a negative outcome than changing it. Extending and replicating the research of Kahneman and Tversky (1982),

Landman (1987) finds that experienced regret is higher after action than after inaction. Additionally, she predicts that the experience of pleasure is higher for positive outcomes when they result from an active decision.

However, even if regret seems a reasonable explanation for the decision-maker's behavior, regret may not (fully) explain the second party's punishment decision (DeScioli et al. 2011b, p. 442). A recipient may not feel regret himself as he cannot determine the outcome as this has already been implemented by the decision-makers. One possibility may be that the recipient empathize with the decision-maker and punish less, as he anticipates that the decision-maker chooses the default to minimize his own regret.

Recommended Option and Inertia

A default option given by policymakers or a company can be considered the recommended and appropriate choice. Using a hypothetical questionnaire study, McKenzie et al. (2006) highlight that most participants chose the default option because they thought the experimenter wanted them to do so. The appropriateness and thus acceptability of a selfish default or status quo option is higher than for an option that was not declared as default or status quo (pp. 414-415).

Changing the status quo or default option takes time and effort, resulting in inertia to stick with the presented or already implemented option. Procrastination to save time searching for alternatives is common and may be a major reason for the status quo bias and default effect (Thaler and Sunstein 2003, p. 177). Carroll et al. (2009) exhibit that the active decision to enroll in a savings plan is higher when employees have a fixed deadline. However, they are often reluctant when the default is an opt-in decision that requires them to actively enroll (Madrian and Shea 2001, pp. 1177-1181).

Maintaining a Good Self-Image

A predominant motive for donating is public awareness, or the demonstration of generosity, as completely anonymous donations are rather rare. Many real world examples and experimental results show how people's behavior is influenced by an intrinsic motivation to help, which diminishes (or even disappears) when they are rewarded for helping. For donations (see e.g. Frey 1997; Gneezy and Rustichini 2000) or volunteer work supply (see e.g. Frey and Götte 1999) this has already been documented in the (experimental) literature. These studies explain that people care about their self-image and want to be seen in a good light by themselves and others (Bénabou and Tirole 2006, p. 1653).

The widespread result of the dictator game, in which people voluntarily give money to anonymous individuals even if they do not see them again, cannot be explained by fairness considerations alone. Andreoni and Bernheim (2009) provide a different explanation, whereupon people want to be seen as fair and care about their self-image. Their theoretical prediction incorporates the audience (the experimenter in a laboratory experiment) into the utility function. This explains why people use the 50-50 norm, according to which they divide the endowment equally between themselves and their partner, even when they can give less (or more) (pp. 1607-1611). Using a dictator game, they confirm their theoretical analysis and highlight that "people are fair-minded to varying degrees [... and ...] like others to see them as fair" (Andreoni and Bernheim 2009, p. 1624). As the experiment by Dana et al. (2007) reveals, people often take advantage of the 'moral-wiggle-room', where decision-makers stay ignorant and hide their selfish decisions. This reinforces the importance of the self-image as people avoid being seen as unfair and preserve their "illusory preference for fairness" (p. 67). The same is true for omissions. Since the intention is less obvious, people may prefer omissions to maintain a positive (self-)image. In this way, individuals "construct a justification of their behavior consistent with their knowledge of the situation" (Vaal 1996, p. 163). Table 3.1 summarizes the previously presented explanations of why people may choose the default.

Table 3.1.: Explanations for the Default Effect and Omission Bias

Explanation	Description
Rationality	Choosing the default is rational if uncertainty and costs of switching are higher than expected gains.
Loss aversion	Outcomes by acts of omission are always perceived better than the same outcome by commission.
Norm theory	Acts of commission are evaluated conditional on the resulting outcome as a result of which unpleasant outcomes are seen as more worse/negative than outcomes by omission.
Regret theory	Choosing the default and acts of omission can reduce the feeling of regret if a negative outcome results.
Recommended option and inertia	A default can be seen as the recommended and appropriate option. Changing it may involve effort which may be avoided as people are inert.
Maintaining a good self-image	People want to be seen as fair (by themselves and others) so they prefer an act of omission where the underlying intention is less obvious.

3.2. Social Norms

In most societies, individual behavior is influenced by social norms, as they constrain self-interest and lead to more prosocial and other-regarding behavior. To help others, people voluntarily forgo material gains (Bicchieri et al. 2022, p. 59). The decision-maker's behavior is therefore often motivated by the wish to comply with (social) norms. Norms are social because they are based on shared beliefs and specify how a person should behave in a specific situation. Therefore, social norms differ depending on the situational context leading to different behavior in different situations. When individuals diverge from the expected behavior, and thus violate the social norm, a form of penalty is very likely, either as a feeling of embarrassment, guilt, shame or as monetary losses. In contrast, norm conformity is often associated with positive emotions (Elster 1989, p. 100; Bernhard et al. 2006, p. 217; Fershtman et al. 2012, pp. 140-141). Thus, "acts are judged morally right or wrong depending on whether they obey or violate some moral rule" (Darley and Shultz 1990, p. 529).

Definition 6: Social Norms

Social norms are (moral) rules shared by society, specifying appropriate behavior in specific situations.

According to Fehr and Gächter (2000b) social norms are characterized by three essential features: (1) they are consistent, (2) they have to be shared by groups or members of a group to be social, and (3) informal sanctions enhance norm compliant behavior (p. 160). Krupka and Weber (2013) extend this definition by adding a situational aspect, as social norms refer to actions rather than outcomes (p. 498). When social norms influence individuals, they usually want to obtain other people's approval (or avoid disapproval), so sanctions (such as punishment) are unnecessary. The mere expectation of others' approval and reward encourages norm compliant behavior, which is what makes social norms so powerful. Many people voluntarily comply with social norms because they align with their self-interest (Elster 1989, pp. 104-105; Bernhard et al. 2006, p. 217; Krupka and Weber 2013, p. 499).

When social norms are violated, individuals can expect to be punished, making punishment an important factor in reaching norm compliant behavior. Even in one-shot interactions, where people never interact again, costly punishment frequently occurs (for a survey, see Chaudhuri 2010). Punishment is often distinguished by the form of sanctions, direct vs. indirect punishment, and by the person or institution carrying it out, second vs. third-party punishment (Fehr and Fischbacher 2004, p. 63). Direct punishment means that individuals

(or groups) have to pay for the sanction, either as a material cost (time/effort) or as a personal cost (payoff reduction). Often, individuals are willing to pay for the possibility of punishment, even though they do not gain materially from it (= altruistic punishment), exposing themselves to retaliation. In contrast, indirect punishment often does not come at a direct cost as it implies retaining rewards for the norm-violating person (Balafoutas et al. 2014, p. 15924).

Comparing direct and indirect punishment in a field experiment, Balafoutas et al. (2014) highlight that individuals prefer indirect punishment (withholding reward) over direct punishment, especially when they receive no reward for direct punishment (p. 15926). Even in the absence of material gains, punishment by an unaffected observer is a common finding in many experimental studies (see e.g. Fehr and Fischbacher 2004; Bernhard et al. 2006; Henrich et al. 2006). Fehr and Fischbacher (2004) compare second- to third-party punishment and exhibit that the person directly affected by the norm violation punishes more than an unaffected observer. In this context, they argue that strong (negative) emotions are the key driver of punishment because they induce a strong dis-utility for individuals (pp. 84-85). Negative emotions, such as anger, are also identified by Fehr and Gächter (2002) as a main driver for (altruistic) punishment (p. 139). Apart from functioning as a sanctioning mechanism, punishment has a strong influence on cooperative behavior in the public goods game, gift exchange game, or trust game (an overview of related studies is presented in Fehr and Fischbacher 2004). Relating to this, Fehr and Gintis (2007) specify that some people seem to have internalized norms of cooperative behavior. These people are strongly reciprocal and cooperate on the condition that others cooperate, while they punish non-cooperative behavior, even if it comes at a cost to themselves (p. 49). However, as cooperative behavior is not part of this study, it is not further elaborated here.

Based on these aspects, we can formulate our next working hypothesis regarding the recipient's beliefs about how decision-makers may decide. If the decision-maker does not act as the appropriate norm postulates, punishment is likely to occur. On the other hand, if a receiver would act in the same way as the decision maker, we expect less punishment because no applicable norm is violated for him.

Working Hypothesis 4 (Beliefs and Expectations) *We expect less punishment if recipients would decide in the same way as decision-makers do.*

Unfortunately, there are only a few studies explicitly dealing with gender differences in punishment, and the few results that do exist are challenging.[16] In a standard public good

[16]Of course, in almost every experiment, data regarding gender is collected. However, even if gender-specific

game with punishment, Burnham (2018) finds no significant punishment differences between men and women, although women punish less than men in absolute terms (1.76 to 2.46 out of 30 possible punishment points). However, when monetary incentives for a higher rank (based on payoff achieved during the session) are implemented, men punish significantly more. Women assign an average of 4.34 punishment points per round, while men assign twice as many punishment points per round (8.74, t-test, $p > 0.01$) (pp. 3-4). Singer et al. (2006) show that men's brains respond with less empathy when they observe a person suffering pain in a sequential prisoner's dilemma, while at the same time the reward-related areas of their brains show more activity. Consequently, men emphasize with fair-acting individuals while preferring physical punishment for unfair-acting individuals.

Although results on differences in punishment between men and women are rare and should be taken with caution, in conjunction with the finding that men are less generous than women (among others, see e.g. Eckel and Grossman 1996; Selten and Ockenfels 1998), we state the following working hypothesis about women's punishment behavior.

Working Hypothesis 5 (Genderspecific Punishment Differences) *We expect women to punish (slightly) less intense and less often than men.*

The experience of guilt is another reason why people behave pro-socially even without punishment. In this context, social norms are moral expectations, because they can shape the recipient's beliefs of what constitutes appropriate behavior. These may be anticipated and expected by the decision-maker, who follows the norm to avoid guilt (Charness and Dufwenberg 2006, pp. 1595-1596). As social norms specify the appropriate behavior and thus constitute a kind of social rule that should be followed, immoral behavior is the act itself that (potentially) violates the norm (Elster 1989, p. 100; Darley and Shultz 1990, p. 529). An individual feels guilty when he hurts someone – for example, when he chooses an unfavorable allocation and believes that the other person would have expected a different behavior. "Guilt averse players hold beliefs about the expectations of others (second-order beliefs) and, thus, hold beliefs about the potential let-down their behavior might cause" (Bellemare et al. 2017, p. 233). Consequently, people try to avoid feeling guilty because they are guilt averse. More specifically, the more the helper believes that the victim expects him to help, the more likely the victim is to receive help in distress. If the norm is to help someone in need, the victim's expectations are influenced by this norm, which in turn influences the

differences are evaluated and reported, statements about general differences in punishment behavior are rare.

helper's behavior, who would otherwise feel guilty.[17] Additionally, to prevent feelings of guilt after a selfish or payoff-maximizing decision, an individual may apply a social norm that is consistent with his selfish behavior. In this way, selfish behavior is justified, so that he does not need to feel 'bad' (Fershtman et al. 2012, p. 141). Since punishment does not lead to payoff maximization in our setting, this aspect of guilt aversion is not applicable.

In this study, transferring guilt aversion to recipients' punishment behavior may imply that a recipient does (not punish) because he believes that the norm is (not) to punish. For example, given an equal outcome, the other group members, decision-maker and (or) recipients, do not expect him to punish because the resources are already equally distributed. Nevertheless, a recipient may want to punish for his own reasons, but does not do so in order to avoid feeling guilty. In contrast, when the unequal outcome is implemented, a recipient may believe the other recipients expect him to punish, even though he does not want to. In this case, he may either feel guilty for not meeting the others' anticipated expectations or punish even though he does not want to punish.

A persons' (group) identity and social norms are closely related, as each identity is accompanied with specific group norms.[18] Whenever an individual identifies with a group, he accepts the group norms (Fershtman et al. 2012, p. 142). If individuals are identified with their group, they follow the group norms because they want to conform with the group-specific expectations and avoid behavior that signals disloyalty to the group. Thus, belonging to a group can lead to norm conforming behavior (see e.g. Akerlof and Kranton 2000; Chen and Li 2009).

In a dictator game with 850 participants, Gächter et al. (2017) examine the effect of peers on individuals' perception of norms, especially norms of fair allocations. By varying the decision situation in two ways (give vs. take option and peer vs. no-peer), they first measure the perception of what is considered socially appropriate. Second, they measure how those perceptions translate into actual behavior. Overall, they find heterogeneity in their experiments, where the presence or absence of peers strongly influences appropriateness. Half of the dictators choose the same transfer level regardless of the recipients' wealth, and a third decrease their transfers when the recipient's wealth is higher (Gächter et al. 2017, pp. 73-81). Bicchieri et al. (2022) analyze whether social distance from others mitigate norm erosion after a norm violation. In a laboratory experiment, individuals can give to or take from a

[17] See Battigalli and Dufwenberg (2007) for a general theory of guilt aversion based on a psychological game theory approach.

[18] The relationship between norms and group identity is further addressed in Section 4.1.2.

charity and (not) observe what others did in previous rounds (pp. 59-60). In sum, observing others taking reduces donations by about one fifth, while observing giving does not increase donations. Observing inappropriate behavior apparently leads to norm erosion, which cannot be compensated for by observing appropriate behavior. However, a lower social distance from others reinforce norm compliant behavior. Even in the absence of formal sanctions group identification leads to more adherence to norms, highlighting the importance of the social context (Bicchieri et al. 2022, pp. 64-67).

These results emphasize the situational context of social norms, where the appropriate norm and justified behavior depend on the situation.

Summary 3.5: Social Norms

As moral rules, social norms specify how individuals should behave in specific situations. Depending on the situational context, they specify how resources should be divided fairly. Deviations from expected behavior, thus violating a social norm, often leads to direct or indirect sanctions.
Identity and norms are correlated in two directions. First, norm conformity increases with a shared group identity, reinforcing norm-compliant behavior. Moreover, a shared group identity can shape (or override) one's own norms through shared group norms. Second, group identification is promoted when other group members conform to the norms.

3.3. Preferences

In economics, the individual is usually classified as rational and self-interested, maximizing his utility. However, empirical findings such as giving behavior in the dictator game (among others, see e.g. Bardsley 2008; Engel 2011) or contributions to public goods (among others, see e.g. Isaac et al. 1984; Chaudhuri 2010) challenge this self-centered view, as results cannot be explained by rationality and self-interest alone. More specifically, justice considerations about how resources should be fairly distributed in a society, as well as individual preferences, come into focus.

In this section, we review the theoretical and empirical literature on preferences and fairness, or rather, justice considerations in order to clarify the related expectations for punishment behavior in our study. Therefore, we first focus on preferences and distinguish four

types of preferences: time, risk, social, and distributional preferences, focusing on the latter as they are most relevant to our setting.

The literature does not always make a clear distinction between distributional and social preferences. In this study, we take distributive justice as defining different criteria or principles according to which a society could (or should) distribute resources. Of course, as we will see later, these criteria often overlap with social preferences. However, as, for example Konow and Schwettmann (Konow and Schwettmann 2016), specify, "we interpret justice as being about distributive moral preferences [...], we leave aside unconditional altruism [...], and reciprocity, i.e., preferences to reward kindness or punish unkindness" (p. 84). According to this definition, pure altruism and reciprocity are not part of distributional preferences. However, they are covered by social preferences. Similarly, Croson and Konow (2009) distinguish two types of social preferences: distributive preferences about outcomes or end states and reciprocal preferences about intentions or player types (p. 4).

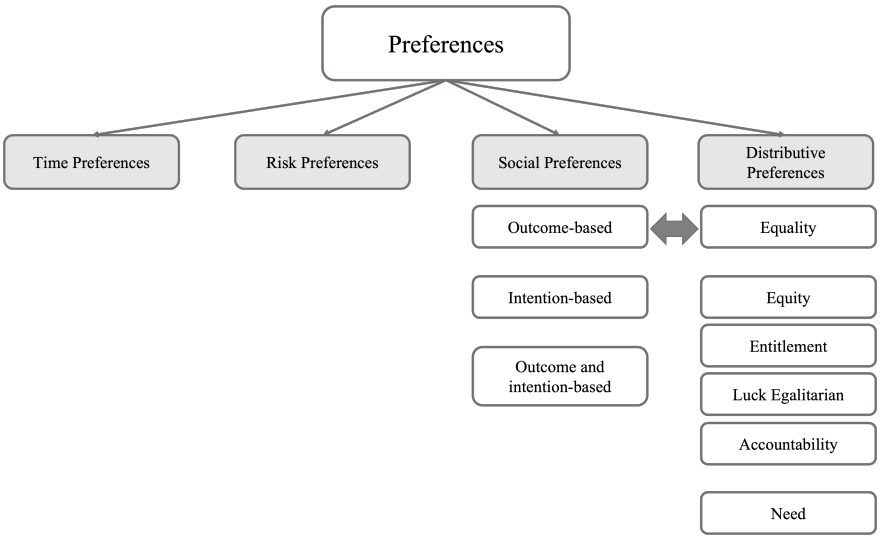


Figure 3.2.: Four Types of Preferences

Figure 3.2 illustrates the four types of preferences with the related models and criteria, with distributional criteria already classified according to their similarity. In the following sections, these models, respectively criteria, are presented in detail from a theoretical and experimental perspective. Much of the research attempts to decipher which preference influences individual behavior and to what extent. The models and approaches presented here form the basis for the punishment motives in Section 6.5 and are considered as control variables in the regression analysis. In choices involving a trade off between a smaller amount now or a larger amount later, a vast body of literature has shown that people have preferences that

are biased towards the present (among others, see e.g. Laibson 1997; O'Donoghue and Rabin 1999). However, as we do not have a time component in this sense in our study, we leave them here aside and focus on social preferences (Section 3.3.1), risk preferences (Section 3.3.2), and distributional preferences (Section 3.3.3).

3.3.1. Economic Models of Social Preferences

Usually, people are classified as acting rational and self-regarding, who want to maximize their own payoff and care only about other people if they impact their own payoff. However, experimental and empirical findings prove that people are not as self-regarding and rational as predicted by theory (Fehr and Gintis 2007, p. 45). Rather, they exhibit social preferences[19] insofar as the payoff of others is also part of their utility function. As punishment in this study is costly and does not bring any material benefit, social preferences may explain recipients' punishment behavior (Akbaş et al. 2019).

To "explore the economic consequences of social preferences, wherein agents have preferences that are measured over their own and others' material payoffs" (Fershtman et al. 2012, p. 131) can help to explain economic phenomena. However, conclusions should be drawn with caution, as the influence of social preferences can be distorted by self-interest, especially in strategic environments. A proposer's generous offer in the ultimatum game can be seen as a preference for equal outcomes – even if it is actually pure self-interest, as small offers are often rejected, which the proposer may anticipate (Croson and Konow 2009, p. 7). Models that receive much attention focus on altruism or fairness, such as in the outcome-based models by Fehr and Schmidt (FS) (1999) and the Equity, Reciprocity, and Competition (ERC) model by Bolton and Ockenfels (2000), where agents are inequity averse and prefer equal payoffs.

Fehr and Schmidt (1999) model fairness as self-centered inequity aversion, meaning that people are interested in getting as close as possible to an egalitarian distribution. A sharp deviation from an equal payoff distribution is associated with negative utility. Thus, subjective utility differs depending on the relative standing of others, who are either ahead or behind. Another outcome-based model of social preferences is the ERC model by Bolton and Ockenfels (2000), in which the payoff and relative position compared to other players are of fundamental importance.

[19] Social norms and social preferences are closely related and sometimes used similarly, although they differ in terms of their stability. Social preferences are stable in different situations, while social norms are situation-dependent and can change from one situation to another (Fershtman et al. 2012, p. 142).

Working Hypothesis 6 (Outcome) *According to outcome-based models, punishment occurs whenever payoffs are not equally distributed. Decision-makers choose the equal allocation if they are inequality averse, and recipients punish an unequal outcome (regardless of the individual choice) even though it is costly.*

Behavior that selfish or purely outcome-oriented preferences cannot explain is reciprocal behavior. "In response to an act of party A that is favorable for party B, B is willing to take costly actions to return at least part of the favor (positive reciprocity), and in response to an act that is perceived as harmful by B, B is willing to take costly actions to reduce A's material payoff (negative reciprocity)" (Falk et al. 2003, p. 20). Accordingly, reciprocity means to reward kind acts and punish unkind behavior, emphasizing the importance of the intention behind the action.

A theoretical model that implements intentions and explicitly controls for reciprocity is developed by Rabin (1993). Relating to this, an action is considered fair if the intention behind the action is kind, whereas an unfair action implies an unkind intention. The perceived kindness or unkindness of an action is determined by the status quo or reference point, such as the payoff distribution or a default. The theory of sequential reciprocity developed by Dufwenberg and Kirchsteiger (2004) is an extension of Rabin's concept including strategic situations with a sequential structure. The relationship between the actual action to the possible alternatives and beliefs is essential in theories of intention and reciprocity. Thus, the affected party, such as the recipient in an ultimatum game, evaluates the intention and fairness of an action as a function of possible alternatives, while the decision-maker takes the beliefs and expectations of the recipient into account before making his decision. Here again, counterfactual thinking takes place.

Similar to Rabin (1993), Levine (1998) provides a theoretical analysis of the ultimatum game. According to this, people behave generously, i.e., altruistically, provided they expect others to be generous or altruistic toward them as well. If the proposer gives an equal or fairer share to the recipient, he hopes that this is appreciated by accepting his offer. The decision is thus seen as a signal of his intention or altruism (pp. 593-595). Consequently, generous behavior by the decision-makers may lead to generous behavior by recipients, such as lower rejection rates in the ultimatum game or less (no) punishment. Recipients may have internalized the 50-50 norm and expect decision-makers to choose the equal allocation as it is the appropriate behavior.

Although many theoretical models and experimental results focus on the decision-maker's actions, they also help to predict the recipient's punishment behavior. For example, they address differences in reciprocal responses as a function of other available alternatives a

decision-maker can choose (see e.g. Brandts and Solà 2001; Falk et al. 2003; Sobel 2005). Falk et al. (2003) show that rejection rates for identical offers in the ultimatum game differ, depending on other possible offers. A (8,2)-allocation is more likely to be rejected if the other allocation is equal (5,5) in comparison to an unequal allocation (2,8) or a very unequal allocation (10,0) (44.4% versus 26.7%, 8.9% respectively), which is significantly different (Cochran Q-test, $p<0.0001$). Thus, when the proposer has the opportunity to choose a more equitable offer, an unequal offer is more likely to be rejected. However, rejection rates are much lower if the proposer can only choose between two unequal offers. Choosing an unequal offer when an equal offer is available signals a rather unfriendly intention on the part of the proposer (pp. 20-24). As already highlighted in Section 2.1.3, by evaluating an outcome or action in comparison to other possible alternatives, a different reality is mentally constructed, which is then used as a reference point. This form of counterfactual thinking determines the perceived cause of an (unpleasant) outcome and therefore explains the differences in punishment behavior (Mandel and Lehman 1996, p. 450; Wells and Gavanski 1989, pp. 161-167). Using the direct response method, Cox and Deck (2005) cannot confirm these results as their rejection rates are generally lower. However, repeating the experiment using the strategy method as in Falk et al. (2003) leads to the same results (pp. 627-628). A general discussion comparing the two experimental methods is covered in Section 6.4.

Another way to examine the role of intentions is to compare choices made by people with choices given by nature (among others, see e.g. Blount 1995; Offerman 2002). People tend to evaluate the same outcome differently when it is determined by nature, as in a random draw, or when it is chosen by a person. Blount (1995) examines the interaction of absolute payoffs and a relative fairness component that takes into account how the outcome is reached. Using ultimatum games, which require students to indicate the lowest amount they are willing to accept as a recipient, the division of the endowment is either determined randomly, by a proposer, or by a neutral third party. She points out that rejection rates are significantly lower when the allocation is determined randomly (\$1.2) than when it is made by a human (\$2.91 for a proposer and \$2.08 for a neutral third party, respectively). Applying a Kruskal-Wallis ranking test shows that differences are significant ($\chi^2 = 7.43, p < 0.025$).[20] Consequently, the conclusion that people care not only about absolute payoffs and final outcomes but also about fairness and how the outcome is reached is quite logical (Blount 1995, pp. 131-132). Applying intention-based models to this study, punishment occurs even if it does not reduce the payoff differences between participants, which we specify in our next working hypothesis.

[20]The theoretical approach of Sebald (2010) allows to explain these experimental findings.

Working Hypothesis 7 (Intention) *An unkind action, such as choosing the unequal allocation, is punished because it reveals an unfriendly intention. In contrast, when the equal allocation is chosen, models of intention-based reciprocity do not predict punishment.*

A theory combining the intention of an action and its consequences has been put forward by Falk and Fischbacher (2006). They propose a framework in which the perceived kindness of an action depends on the resulting outcome and the underlying intention (p. 294). In examining the prediction of the various fairness models, Falk et al. (2008) support the relevance of the intention-based and outcome-based approach, as the combination of both is the best suited to predict punishment behavior.

Working Hypothesis 8 (Outcome and Intention) *Relating to this, outcome and intention matters so that an unequal choice is punished if the unequal allocation results because it implies an unfriendly intention.*

A different perspective on kindness, or how exactly to define kindness, is highlighted by Çelen et al. (2017). They propose a definition of kindness based on the concept of blame. The unkindness of an action depends on whether the recipient would act in the same way as the decision-maker. Thus, a decision is judged according to subjective criteria and the situational context. Consequently, each person evaluates the kindness or unkindness of an action differently. If he acts nicer than the decision-maker, he is willing to blame him for his action (pp. 62-64). This perspective focuses not only on the outcome, but also on the recipient's beliefs, or expectations, about how a decision maker might decide. We have already covered this aspect in the previous section, when we specified Working Hypothesis 4.

Counterfactual thinking, or thinking in terms of possible alternatives, is also relevant here, considering the context and the action that leads to an outcome. Further experimental evidence supports the impact of the decision situation on individual reactions. From a rational perspective, the same proportion of giving/taking amounts are expected in a dictator game. Even if the outcome in a dictator game with a give or take option is the same, taking money away is often seen as more unfriendly and unkind than not giving. Additionally, to take money away may induce other social norms, so the reciprocal reactions differ, again highlighting the context-dependent aspect of social norms (among others, see e.g. List 2007; Bardsley 2008; Ramalingam et al. 2019). For example, Bardsley (2008) shows that generosity reverses and more unselfish behavior occurs when people can take money, which is inconsistent with models of social preferences (pp. 124 -128). List (2007) also illustrates that giving rates decrease when taking money is possible, and Ramalingam et al. (2019) exhibit that

taking money is punished more than less unkind behavior, such as not giving, even if the payoff is the same in both settings. In a principal-agent setting, Gurdal et al. (2013) find that it is the agent's decision that matters, even if the agent is not responsible for it. Counterfactual thinking explains this punishment behavior as people compare the actual outcome with possible other outcomes that could have been achieved if the decision-maker had acted differently (pp. 1207-1210).

In their pertinent paper, Andreoni and Miller (2002) investigate whether altruism is explainable by self-reflexive utility-maximization and is thus (entirely) rational. In a modified dictator game, where different endowments and payoffs are presented in a randomized order, they highlight that 98% of all subjects maximize their utility. Only a quarter are purely selfish, implying that the reminder exhibit varying degrees of altruism, while 14 % choose equal payoffs. Consequently, fairness preferences are heterogeneous and people differ in their preference for fairness. Some do not care about fairness, while others' perceptions range from equal shares to pure selfishness (pp. 737-745).

Summary 3.6: Social Preferences

So far, these results exhibit a clear picture: people are not only interested in the end result, as assumed by the inequality aversion model. Rather, they also consider the intention behind the action, as predicted by intention-based models. However, the influence of outcome-based models should not be underestimated. A combination of both aspects, outcome and intention, seems to be the best approach to explain individual behavior.

3.3.2. Risk Preferences

Risk preferences are typically measured with lotteries that have different expected values and degrees of risk. Based on individual choices between these lotteries, a risk aversion parameter is estimated (Erkut and Reuben 2019, p. 40). Holt and Laury (2002) model how to measure risk aversion in experimental settings and investigate whether the degree of risk aversion varies with the amount of payoffs. In a paired lottery framing, participants have to choose in 10 situations between two lotteries A and B, while payoffs in lotteries A are less variable than lotteries B. Summarized, even when payoffs are low, two-thirds of participants are classified as risk averse, while risk aversion increases when real payoffs increase, as more decisions are made for the safe option. However, as the risk attitude is not the main research interest in

our experiment and we measured the risk-attitude self-stated on a five-point likert scale, we focus mainly on the implications of risk-aversion on the individual behavior.

The finding that women are more risk-averse that men is robust in many experimental settings (among others, see e.g. Holt and Laury 2002; Fehr-Duda et al. 2006; Croson and Gneezy 2009; Eriksson and Simpson 2010; Charness and Gneezy 2012). Eckel and Grossman (2008) provide an overview of findings from economics, for a review of psychological research on gender differences see Eagly (1995).[21] For example, in financial risk taking Fehr-Duda et al. (2006) examine in a laboratory experiment whether men and women weight the probabilities with which outcomes result differently. They find out that women underestimate larger probabilities more than men do, especially in an investment frame compared to an abstract framing. Although gender differences are not their primary interest, Holt and Laury (2002) exhibit that men are significantly more risk averse than women ($p < 0.05$) for low payoffs, which disappears for high payoffs (p. 1651).

There are a bunch of possible explanations as to why women are more risk averse than men. The emotional reaction to uncertainty differ between the gender so that women behave differently in risky situations. Additionally, men are often classified as more confident and more prone to challenges so that they may evaluate the winning/loosing probabilities differently (Croson and Gneezy 2009, p. 454).

For our study, the implications of different risk attitudes, especially concerning the punishment behavior, are of special interest. More specifically, we are interested in whether risk-averse participants generally punish less (more) than risk-affine participants. Additionally, the conclusion that women are more risk-averse than men may appropriate, so that gender specific punishment differences may also result from different risk-attitudes. However, to our knowledge, there is no experiment that addresses punishment behavior as a function of risk attitude, so we can only speculate about the direction of effect here. Since a default reduces uncertainty, potentially influencing preferences (Dhingra et al. 2012, pp. 69-70), and changing a default option is associated with more uncertainty and risk (Samuelson and Zeckhauser 1988, pp. 33-35; Johnson and Goldstein 2003, p. 1338), we state the following working hypothesis.

Working Hypothesis 9 (Risk Attitude) *We may expect risk-averse people to punish less than risk-affine people.*

[21] However, recently, Nelson (2014) claims three reasons why published results on gender differences on risk aversion have to be taken with caution. First, generalizations have often been taken from mean differences that do not specify the degree of differences. Second, other confounding factors, like cultural aspects, are not always captured. Third, often studies that do not find significantly gender differences are not published, so a publication bias may lead to more studies with significant differences.

3.3.3. Distributive Preferences

When goods, burdens, or resources are distributed in a society, social inequalities may occur. Objective and subjective criteria are relevant to the evaluation of inequalities, with subjective evaluation depending strongly on individually applied norms and preferences, so that people often evaluate the same social inequalities differently (Hülle et al. 2018, pp. 663-664). Distributional preferences are contextual and pluralistic, with context determined by both institutional factors and individual characteristics. Regarding the pluralistic argument, Konow (2000) considers "that most peoples' values [on fairness] may be accounted for by several fairly simple principles, any of which may dominate depending on the context. [...] People may weight competing fairness principles differently or may perceive and evaluate the factors relevant to even a single principle differently" (p. 1073).

Apart from the long-established principles of equality and equity, need, entitlement, and accountability have been incorporated to the debate on redistribution. The recent political and social debate supports them by focusing on individual characteristics and controllable factors. There is a large amount of research highlighting that fairness considerations depend on perceptions of individual's accountability, i.e., factors that can (not) be controlled (among others, see e.g. Konow 2000; Krawczyk 2010; Mollerstrom et al. 2015). Relating to this study, accountability, which is synonymous with responsibility (Schlenker et al. 1994, p. 632) and luck egalitarianism are relevant. The latter argues that an individual is to be compensated for outcomes he is not accountable for (among others, see e.g. Dworkin 1981; Knight 2009; Cohen 2011).

Roots of Distributive Justice

Today, many disciplines such as philosophy, psychology, sociology, political sciences, and economics deal with the issue of distributive justice (Traub 2020, p. 2). In (political) philosophy it has a long tradition and goes back to Aristotle and Aquinas as well as other philosophers. Speaking of distributive justice often means the distribution of physical goods, whereas social justice is less restricted in that it rather pertains to the way in which access to and distribution of resources, goods, and burdens takes places (Miller 2001, p. 2; Siebel and Schramme 2020, pp. 21-22). In this context, Miller defines social justice as "how the good and bad things in life should be distributed among the members of a human society" (Miller 2001, p. 1). The concept of justice is also applicable to the right action and is thus related to individual morality evaluating individual's actions as just or unjust. In political philosophy, however, the focus is not on the individual, but on institutions that are considered just or

unjust (Siebel and Schramme 2020, p. 22).

Although distributive justice has a long tradition in political science and philosophy, it did not receive much attention in economics in the twentieth century. Today, however, it "occupies a prominent place in theoretical and empirical economic research" (Konow and Schwettmann 2016, p. 83). Frohlich and Oppenheimer (1994) distinguish two lines of theoretical literature, some of which sometimes overlap. First, there is the literature on impartial reasoning, where Rawls' prominent theory of justice can be categorized. It determines standards for a fair distribution of income under idealized conditions requiring individuals to agree on those standards. Second, the behavioral literature views income distribution as choices among social insurance programs involving risks, losses, and uncertainty (pp. 147-148).

Economic justice research can generally be divided into two dimensions, which are (mostly) conducted independently of each other. On the one hand, there is a prescriptive and mostly theoretical dimension that addresses the normative question of what 'ought to be'. On the other hand, there is a steadily growing descriptive literature that captures what 'is', including mainly experimental (and empirical) findings that impact the descriptive part of the literature (Konow and Schwettmann 2016, pp. 83-84). In this section, we begin with presenting theoretical dimensions of different justice criteria, while their empirical verification, or rather what criterion is actually applied, follows afterwards.

Standard Principles of Distributive Justice

Numerous theoretical and empirical studies have addressed the normative question of which principles of distributive justice to apply, i.e., how resources or goods *should* be distributed. In economics, achieving efficiency is often the main objective in the allocation of resources. There are various concepts of efficiency, the best known being the Pareto principle (Pareto 1906). A payoff is Pareto efficient if at least one person is made better off while no person is made worse off (Konow and Schwettmann 2016, pp. 84-87). In our experimental setup, the Pareto principle is not satisfied because switching from the equal (unequal) allocation to the unequal (equal) allocation only makes one party better off, while the other party receives less than without switching. Consequently, the default option can be retained even if it is perceived as unfair (as in the case of the unequal default option) and a switch would lead to a fairer distribution.

The best-known distributive principles are equality and equity, which are very different from each other and (usually) mutually exclusive (Traub 2020, p. 2).[22] The equality principle, the oldest criterion of justice, states that everyone (should) receive the same amount of money, goods, or resources. Additionally, one's material well-being as well as inequality aversion, in the sense of disutility when the allocation deviates from an equal distribution, are also embodied (Konow and Schwettmann 2016, pp. 84-85). Therefore, it aligns with the predictions of inequality aversion, according to which payoff differences should be minimal. Combining equality with efficiency yields the maximin-rule, according to which the minimum payoff in the group should be maximized (Engelmann and Strobel 2004, pp. 857-858).

In several one-shot distribution experiments, Engelmann and Strobel (2004) compare the relevance of efficiency considerations, maximin preferences and inequality aversion to examine the relative performance of the FS and ERC model in predicting distributive decisions. They highlight that efficiency and maximin preferences have a strong influence on distributive decisions, while inequality aversion is less influential. When maximin preferences are neglected, the FS models predicts behavior better than the ERC model (pp. 860-866).

In equity theory, based on the early work of e.g. Homans (1958) and Adams (1963), allocations should be based on the individual's current contribution or effort, with everyone receiving the share proportional to their input. Consequently, each individual's input-output relation is the same (Becker 2012; Hülle et al. 2018). However, it is far from clear what criteria are used to determine the proportionally of allocations. In desert theories, for example, people are held responsible for inequalities resulting from justifying attributes such as effort, while unjustified attributes such as luck are not considered responsible for (Konow and Schwettmann 2016, p. 91). Gill and Stone (2015) develop a theoretical framework in which people claim the amount they feel they deserve. Therefore, they consider a reference point that depends on comparisons with others and that is mutually agreed on. Accordingly, someone who works more deserves more and must receive more payout to be satisfied. If he works more and receives less than someone who works less, he is dissatisfied. Respectively, someone who works less deserves less and should receive less. If he receives more, even though he works less, a feeling of guilt may result. As people are loss averse around the reference point, receiving less with more work is more unpleasant than receiving more with less work (pp. 42-45).

People with a better final grade may be viewed as more deserving because they have worked harder to achieve a better grade. Unfortunately, the exact workload is not measurable in our study, so this notion of desert cannot be calculated for the latter analysis.

[22] Except for a case in which all parties contribute the same share resulting in an equal distribution.

Although a higher grade does not automatically mean more effort, the idea of deservingness as an indicator of how much someone wants is worth noting.[23]

Recently there has been a growing literature focusing on need as a third relevant criterion for distributional decisions. Individual need is defined as the amount of goods or resources necessary to prevent the individual from harm (Miller 1999).[24]

Hülle et al. (2018) further define entitlement as a relevant criterion specifying that "benefits and burdens should be allocated on the basis of specific entitlements that are themselves based on ascribed characteristics (e.g., social origin, sex) or on status characteristics that have been acquired in the past (e.g., occupational status)" (p. 668). Thus, distributions should not depend on current effort, but on past performance or previously acquired characteristics. To measure individuals' support for the four main distributional principles - equality, equity, need, and entitlement - Hülle et al. (2018) introduce the BSJO scale. As it fits well in the overall context of our study and strengthens the latter analysis, it is also implemented in the questionnaire at the end of our experiment (for more details, see Section 7).

Accountability and Luck Egalitariansism

Two prominent approaches, accountability and luck egalitarianism, distinguish between controllable and uncontrollable factors and address the question of where the responsibility cut is drawn. The accountability principle, based on the work of Konow (2000), "requires that a person's fair allocation (e.g., of income) vary in proportion to the relevant variables that he can influence (e.g., work effort) but not according to those that he cannot reasonably influence (e.g., a physical handicap)" (pp. 1073-1074). Thus, the fair allocation of resources considers controllable factors, such as a person's effort. In contrast, factors beyond an individual's control, such as a disability or pure luck, have no bearing on fairness considerations.

Luck egalitarianism, another egalitarian theory, further emphasizes the role of luck in distributional preferences. The expression luck egalitarian has its roots in the philosophical literature and is mainly associated with the work of Arneson (1989) and Cohen (1989, 2011) whereas the main ideas had already been highlighted by Dworkin (1981), who never called himself a luck egalitarian. Luck egalitarianism distinguishes two forms of luck: option luck

[23] Deservingness can, of course, be interpreted differently in the case of a deserving recipient. As this setting typically involves donations, Engel's (2011) meta-study highlights when the recipient is deserving, fewer dictators give nothing and more than a third give the entire endowment (p. 594). As the recipient is not described as deserving in this study, this research line is not further addressed.

[24] Traub and Kittel (2020) provide an interdisciplinary overview of need-based distributive justice from the perspective of five different disciplines.

and brute luck. Suppose the individual choice leads to good or bad luck. In this case, it is classified as option luck, for which the individual is fully responsible, because he makes a choice and voluntarily accepts a potential risk. Thus, a person does not have to be compensated for option luck. In contrast, brute luck does not involve a conscious individual choice. Consequently, compensation for bad outcomes is appropriate as the individual is not held responsible (Dworkin 1981, p. 293; for a recent discussion of the main ideas as well as the debates that have accompanied luck egalitarian, see Knight 2013).

Bringing together the different approaches, Cappelen et al. (2007, 2010) distinguish four different fairness principles:[25] first, strict egalitarianism where everyone should get the same, so that all factors affecting a person's income production are irrelevant. This principle is consistent with the predictions of the inequality-aversion model by Fehr and Schmidt (1999) and the equality preference highlighted earlier. People dislike unequal income distribution and prefer equal allocations for all. The opposite of strict egalitarianism is libertarianism, in which all people should get what they produce, so all factors affecting production are relevant (p. 431). As the equity principle specifies that the relationship between inputs and outputs should be equal, the two approaches are closely related. The last two principles are rooted in the ethics of equal opportunity. They differ by the exact position of "the responsibility cut, that is for what factors individuals should be held responsible and for what factors individuals should not be held responsible" (Cappelen et al. 2010, p. 431). Both principles, choice egalitarianism and meritocratism, have in common that they hold individuals responsible for some factors but not all. Choice egalitarianism implies that people are not responsible for factors of production that are not under their control. In contrast, meritocratism defines people as responsible only for their personal traits but not for factors unconnected to them (Cappelen et al. 2010, p. 431).

Becker (2012) distinguishes three fairness principles: (1) egalitarianism, where everyone should get the same, (2) meritocratianism, where individual contribution determines the final outcome, regardless of whether it is based on effort or luck, and (3) accountability, which excludes luck from the responsible factors, as it is beyond personal control (p. 690). The first two principles coincide with the classification of Cappelen et al. (2010), whereas the accountability principle is based on Konow (2000).

[25] In an earlier version, Cappelen et al. (2007) only distinguish three principles: strict egalitarianism, liberal egalitarianism and libertarianism. Liberal egalitarianism lies in the middle between the two extreme forms and represents the idea that inequalities arising from factors under individual control should be accepted (p. 818).

Summary 3.7: Accountability and Luck Egalitarianism

As it should be clear by now, accountability and luck egalitarianism are similar in that they emphasize individual's control over an outcome, which determines whether an individual is held accountable and compensated. Both approaches rely on the same underlying norms and do not held individuals responsible for uncontrollable factors, so fair distributions depend only on controllable factors.

In the next section, we present some experimental results concerning the influence of different distributive preferences and fairness considerations, where luck egalitarianism and accountability are not explicitly distinguished. But before turning to that, Table 3.2 summarizes the criteria of distributive justice presented earlier and specify how resources should be distributed by illustrating the appropriate distribution with an example.

Table 3.2.: Distributive Justice Criteria

Principle	Explanation/Definition	Expectations
Equality	Resources should be distributed equally.	Everyone receives the same amount.
Equity	Resources should be distributed proportional to the (individual) input	Someone contributing more gets more, so that the input-output relation is the same.
Need	Resources should be distributed according to need.	A needy person receives more than a non-needy person.
Entitlement	Resources should be distributed according to past performance or previously acquired characteristics.	A person should get more if she worked more in the past and achieved a higher education.
Accountability	People are (not) responsible for (un)controllable factors.	A person should be compensated for factors beyond individual control.

Suppose two people P_1 and P_2 have a side job and work in the same office. Both have similar educational backgrounds and work the same number of hours. However, person P_1 is more efficient and produces more (i.e., writes more letters).

According to the equality principle, both should receive the same amount of money or resources, while the equity specifies that the person producing more (P_1) should get more

because he contributes more. The need principle allocates resources according to (individual) need, so that a needy person should receive more. For example, if person P_1 has the side job just for fun to have more money for vacation or hobbies, while person P_2 does this extra work to have enough money for heating in winter, person P_2 is more needy and should get more. Person P_1 should also get more if he is entitled to receive more, e.g. because of a higher education or acquired skills for the work. According to the accountability principle, a person is not held responsible for factors over which they have no control. If person P_2 has lower productivity, because he has an illness and cannot type as fast as person P_1, he is not responsible for that and should get the same as person P_1. However, if person P_2 has lower productivity because she paints her nail and types slower so as not to ruin her nail polish, she is responsible for her lower productivity and should get less.

3.3.4. Experimental Evidence

This section provides an overview of relevant experimental results in order to identify what might influence the individually assigned responsibility. Therefore, the focus lies on two research lines, which, of course, sometimes overlap and cannot be clearly delineated. The first emphasizes the importance of individual control for (re)distributive behavior. Redistribution is higher when people are not held responsible for their poverty or inequality, as it originated outside their control (among others, see e.g. Fong 2001; Krawczyk 2010; Becker 2012; Mollerstrom et al. 2015). The second line of research focuses on the effect of individual choices in determining the outcome and corresponding fairness views. It has been demonstrated that (re)distributive preferences are sensitive to the way in which endowments are allocated (among others, see e.g. Hoffman and Spitzer 1985; Hoffman et al. 1994; Konow 2000; Cherry et al. 2002).[26]

Usually, experimental results focus primarily on distributive justice and fairness considerations in situations involving some form of production, such as the production of income that is (re)distributed. Although this study does not involve a production stage, the idea behind it can be transferred to this study. The attainment of the graduation and the final grade can be seen as a form of past production. Even though it is not achieved in the laboratory, the mere prior performance is associated with effort, and the conclusion that a higher grade is (typically) associated with higher effort is obvious.

In a dictator game with a proceeding production phase, Cappelen et al. (2010) examine

[26] The experiments by Hoffman and Spitzer (1985) and Hoffman et al. (1994) are further addressed in detail in the next chapter.

which of their four principles suits best. They find that people are not responsible for factors beyond their control (the price) but for factors they can influence (the working time). A strict egalitarian position, where everyone should get the same, is not supported. As a conclusion, the responsibility cut is drawn between personal and impersonal (controllable and uncontrollable) factors (Cappelen et al. 2010). Becker (2012) uses a real-effort task to disentangle the influence of different forms of luck and finds out that redistributive behavior depends on the form of luck in interaction with individual achievement levels (p. 686).

Other experimental studies support these findings. For example, using a dictator game with production, Frohlich et al. (2004) and Konow (2000) analyze fairness considerations and support the distinction between factors under individual control and factors beyond individual control. However, Frohlich et al. (2004) argue for equality when distribution based on accountability is too costly. In a questionnaire study conducted in three different cultures - Belgium, Burkina Faso, and Indonesia - Schokkaert and Devooght (2003) further support the distinction between controllable factors, for which individuals are held responsible, and uncontrollable factors, for which individuals must be compensated. As compensation for need is intertwined with accountability, Bauer et al. (2022) find out that individuals are compensated less if they are seen as accountable for their disadvantage. However, regardless of being accountable, the needier person is always partially compensated for her greater need.

In a risky decision context, Mollerstrom et al. (2015) examine inequality preferences of third-party observers, while either bad option luck or purely random bad brute lack results. Participants can insure against the risk of option luck by paying a fee, while bad brute luck is not insurable (pp. 33-34). Interestingly, they find that uncontrollable bad brute luck is compensated if the individual avoids the risk of controllable option luck. They refer to this new form of fairness principle as choice egalitarianism. It is closely linked to individual choices, regardless of their relevance for the good or bad outcome. In this respect, a choice signals a person's type, i.e., avoiding the risk of option luck or not, which then influences the compensation for bad brute luck (Mollerstrom et al. 2015, pp. 39-40). Similarly, Krawczyk (2010) highlights that the support of redistribution is influenced by the cause of individual poverty, either laziness or bad luck. In their experiment, transfers are higher when winning is randomly determined (brute luck) than due to performance in a task (option luck).

Summary 3.8: Experimental Evidence on Accountability

These experimental results confirm the distinction between controllable and uncontrollable factors in distributional decisions holding people responsible for factors they can control.

There is a large body of experimental evidence on how men and women behave in different experimental games and settings. However, the evidence for gender effects is quite mixed. As it is already addressed in Section 2.2.1 women are (usually) more generous than men. In an ultimatum game where the gender of the proposer is revealed, Eckel and Grossman (2001) and Solnick (2001) find that men receive more than women, while proposer behavior is similar for men and women. Eckel and Grossman (2001) use a face-to-face design revealing that men are less likely to accept low offers than women, while women expect more than men in the anonymous setting of Solnick (2001). Thus, depending on the experimental design, rejection rates in ultimatum games differ between men and women. Giving behavior in a dictator game also differs as a function of anonymity. When the setting is anonymous, women give twice as much as men (Eckel and Grossman 1998), whereas they give slightly more than men when the setting is less anonymous (Bolton and Katok 1995).

Women are also generally more inequality averse than men (Andreoni and Vesterlund 2001; Dufwenberg and Muren 2006) and more loss averse than men (Schmidt and Traub 2002, pp. 245-246). In a modified dictator game Andreoni and Vesterlund (2001) examine gender-specific differences in altruism with different payoff-budgets and relative prices without knowing the gender of the recipient. Overall, there are no significant differences in giving behavior between men and women across all incomes ($2.56 for men and $2.60 for women, $t = 0.24$). However, comparing the different budgets reveals that women give more when prices are high, while men give more when prices are low, suggesting that men are more sensitive to price changes. Additionally, women seem to be concerned about providing equal payoffs, while almost half of men are classifiable as selfish (47.37%), compared to 36.96% of women (pp. 295-301).

Summary 3.9: Gender Differences

Together with previous findings, the presented results suggest that women are more sensitive to the situational context of the decision situation leading to greater behavioral variability as they are more (usually) more risk and loss averse than men.

3.4. Procedural Justice

"Procedural justice refers to the fairness of the methods, mechanisms, and processes used to determine outcomes as opposed to the fairness of the outcomes themselves" (Miller 2001, p. 528). In this respect, distributive justice is often associated with the material outcome, while procedural justice focuses more on the immaterial aspect. Consequently, procedural fairness assumes that the evaluation of a decision is determined not only by the outcome itself, but also by the fairness of the procedure that brings about the outcome. Although the two concepts are in some ways distinct, they are interrelated and they sometimes overlap (Bolton et al. 2005, p. 1054).

Starting with Rawls (1971) and Thibaut and Walker (1975), considerations of procedural justice emerge. Fair procedures help to maintain or reach stable structures in social interactions leading to higher acceptance rates of allocation decisions. The basic idea of procedural fairness is that everyone has an equal opportunity of achieving an outcome, even if the implemented outcome is not necessarily equal (Bolton et al. 2005, p. 1054). This already addresses the main aspect of why procedures are important: fair procedures positively affect the involved person. They can function as a regulation mechanism and increase cooperation (or decrease it if the procedure is highly unfair). When procedures are fair, even unpleasant or slightly unfair outcomes are more likely to be accepted, because people (may) think they deserve less (Cremer and Dijke 2009, pp. 114-115). The process of how an outcome is achieved affects people's satisfaction. As people may feel anger after an unfair treatment (as their moral understanding of the underlying norms is violated) or satisfaction after a fair treatment, emotions influence the fairness judgment. In this context, Folger (1987) finds out that people evaluate an unpleasant outcome as less bad if it comes about through a fair process.

A procedure's fairness is evaluated by comparing the procedure actually applied with the (perceived) fair procedure. The greater the discrepancy between the two is, the less fair is the applied procedure (Vermunt and Steensma 2016, p. 223). Since evaluating whether a procedure is fair or not (or whether it is fair to some degree) requires a reference, counterfactual thinking is relevant to the evaluation. As it has been pointed out previously, the availability of other alternatives and how far they come to mind affect the attribution of responsibility. The same is true for procedural justice, where the applied procedure is compared to other possible procedures that could be considered more just or fair. Consequently, they affect the fairness assessment of the procedure actually applied (Vermunt and Steensma 2016, pp. 226-227).

The concept of utility is usually outcome-oriented, so the influence of procedural fairness is not captured by the standard utility function. Incorporating the concept of procedural fairness into a utility function, Frey et al. (2004) state that people have preferences regarding how an outcome is implemented, which in turn leads to an increase (decrease) in utility. Procedural utility in their sense can be defined as follows: "Procedural utility thus can be defined as the well-being people gain from living and acting under institutionalized processes as they contribute to a positive sense of self, addressing innate needs of autonomy, relatedness, and competence." (Frey et al. 2004, p. 381) Consequently, utility is achieved either through institutions, as they affect well-being, or through feelings in the interaction with others (p. 382).

Summary 3.10: Procedural Justice

People care not only about final outcomes, but also about the process leading to those outcomes. When inequalities are implemented through a fair procedure, they are more likely to be accepted than inequalities through an unfair procedure. Fairness is assessed by comparing the applied procedure with the just procedure.

Using different versions of the ultimatum and battle-of-the-sexes game, in which either payoffs or procedures are biased (unbiased), Bolton et al. (2005) explore the role of procedural fairness experimentally. For example, in one game, a proposer must choose between a favorable (unfavorable) outcome for himself (the recipient), or a lottery that implements one of the two allocations with equal probability. To be paid out, the recipient must accept his choice. Interestingly, the unbiased procedure has the same acceptance rate as the choice of an equal outcome for both players. However, when the random draw is biased with respect to the different probabilities of achieving, unequal outcomes are more likely to be accepted (pp. 1061-1066).

An interesting and closely related experiment by Ku and Salmon (2013) assigns participants' initial positions based on various criteria, such as random, performance, minimal group paradigm, and cooperative behavior. Then, they measure the acceptance of inequalities as well as efficiency concerns between the status positions. Disadvantaged participants must decide about the amount to transfer to the advantaged participants. When the absolute payoff of both is increased, at the same time the inequality between them increases (pp. 113-116). Compared to the random treatment, participants make choices that are not efficient and do not lead to the social optimum. Especially the comparison between random

and performance (as measures with SAT questions) is interesting. It reveals that although participants earn their role, they are not willing to transfer more. The authors explain this with negative emotions, as being worst off may lead to jealousy, envy, or shame, causing transfers to decline. Alternatively, participants may disagree on whether the performance is an acceptable way to assign (dis)advantageous positions. The two other procedures lead to expected behavior insofar as transfer levels are lower compared to randomly assigned roles (pp. 120-126).

In this context, Brock et al. (2013) examine whether fairness considerations are influenced by the comparison of outcomes (ex-post) or opportunities (ex-ante) to achieve those outcomes. Even if the procedure of reaching an outcome is considered fair, for example, if two individuals have an equal chance of winning in a lottery, this does not necessarily lead to a fair outcome because only one of them can win. Using a dictator game with a risky component, they investigate that the ex-ante comparison, i.e., equal opportunities, has a more considerable impact on the fairness evaluation than the outcome (ex-post comparison).

Mertins et al. (2013) examine how participants judge the procedure by which a proposer is selected. The proposer then decides how to allocate the endowment in a group of five, while, the other group members can subsequently oppose this proposal. The perceived (un)fairness of the selection procedure is one factor determining the resistance. As a conclusion, people care not only about the final allocation but also about the process of selecting the proposer.

Working Hypothesis 10 (Procedural Justice) *We expect less (more) punishment if the procedure leading to inequalities is perceived as fair (unfair).*

So far, we have addressed procedural fairness from an economic point of view which is extended in the next chapter by emphasizing legitimate procedures.

4. (Legitimate) Status Differences

Another important construct in this study is status, which implies a ranking of people. A ranking is socially accepted and gives access to resources (Ball et al. 2001, p. 161). Status is often used interchangeably with other constructs such as power or influence, while a clear and distinct definition is rarely used (for a good overview of the inconsistencies and contradictions between scholars and disciplines, see Henrich and Gil-White 2001, pp. 166-167; Blader and Chen 2014, pp. 71-72). Therefore, a conceptualization of status, as well as a distinction from related concepts, is necessary and addressed in the first section. In Section 4.1.1 we start with an overview of essential definitions, where we differentiate status from related concepts, such as power, influence, entitlement, or property rights. In order to have a framework for the following analysis, the section provides conceptual clarity of hierarchy dimensions in society.

With this conceptual clarity in mind, Section 4.1.2 presents status from a psychological and sociological perspective, followed by the economic perspective in Section 4.1.3. Two prominent and relevant psychological theories, namely Status Characteristics Theory (SCT) and Social Identity Theory (SIT), are introduced, and meaningful insights into the importance of status (differences) identified. The economic perspective focuses more on status as part of the utility function and the consequences of status differences for (economic) decisions.

Section 4.2 is devoted to give a theoretical overview of legitimacy. As Tyler (2006) points out, "legitimacy may provide a framework through which actions are evaluated and judged to be just or unjust" (p. 384). First, in Section 4.2.1 two forms of legitimacy are distinguished: (1) normative and (2) empirical. The first concept uses objective criteria that should be met for legitimacy, while the second concept specifies how people evaluate or perceive the legitimacy of an authority or status structure.

In Section 4.2.2, legitimacy is considered from a more general perspective, mainly influenced by political science and the (voluntary) acceptance of legitimate authorities. Therefore, this section already focuses on one important aspect, namely the source of legitimacy, either through distributive justice or procedural justice. How an authority or institution achieves legitimacy is an essential and widely examined question. There are two lines of research: the first considering distributive justice and the second procedural justice as the basis for legitimacy (Crandall and Beasley 2001; Tyler 2006). Crandall and Beasley (2001) propose that legitimacy has its roots in distributive justice. In their view, the outcome, rather than the process that leads to the outcome, is the moral foundation of legitimacy (pp. 83-84).

Tyler and colleagues, on the other hand, state that procedural justice leads to institutional legitimacy insofar as fair procedures increase trust in institutional structures and, thus, their legitimacy (among others, see e.g. Tyler 2001, 2006). As both justice criteria have already been discussed in the previous section, this section complements the presentation and further strengthens the justice criteria' relevance. Since legitimacy is the core element for status structures to be accepted and reach social validity (Bettencourt et al. 2001), Section 4.2.3 brings both concepts together.

In the end, Section 4.3 presents status and legitimacy from an experimental perspective highlighting related experiments on status differences. In these studies, a random group-building mechanism is often compared to a mechanism in which participants earn their status through a task or procedure. As various tasks and procedures are typically used in laboratory experiments, this section concludes with an overview of the most commonly used group-building mechanisms.

4.1. Theoretical Basis of Status

4.1.1. Definition of Status and Related Constructs

The primary purpose of this section is to identify the main types of hierarchy-related constructs and to define them systematically, so they are comparable. Six different constructs are mentioned in the (psychological) literature that distinguishes people or groups. The most important constructs for this study are status, identity, and power, complemented by socioeconomic status (SES), prestige, dominance, and influence. From the economics perspective, entitlement and property right are defined and delineated.[27] Although these constructs share some aspects and are based on the same theoretical foundation, they differ from one another (Blader and Chen 2014, pp. 72-73).

Hierarchy-related Constructs

Ball et al. (2001) define *status* as follows: "A person's status is a ranking in a hierarchy that is socially recognized and typically carries with it the expectation of entitlement to certain resources. There are many hierarchies within which a person might be ranked, based on specific skills or accomplishments or general societal rankings" (p. 161). In addition to this definition, there are others, such as in Chen et al. (2012), where a person's status depends

[27] These definitions are far from comprehensive. There may be other disciplines or researchers who use different definitions. However, for the purpose of this study, the following definitions are sufficient.

on social esteem and worth relative to others (p. 300). However, to bring these definitions together, status has three main components: (1) it is an individual's rank in an entity based on specific skills, underlying characteristics, wealth, existing norms, or ideals, (2) it must be recognized and accepted by the entity, and (3) it is associated to an entitlement or access to resources as well as to favorable handling by others (Ball et al. 2001, p. 161; Henrich and Gil-White 2001, p. 166).

In this respect, an assigned status in society is linked to specific characteristics or skills of the status-holding person, such as his competence, (cognitive) ability, or actions, leading to honor and respect (Weiss and Fershtman 1998, p. 802; Essen and Ranehill 2011, p. 3). Depending on the reference group, however, a ranking is evaluated differently and (can) lead to different status positions (Ball et al. 2001, p. 161). An individual can have various status positions and be part of more than one group simultaneously, as different dimensions induce different rankings (Weiss and Fershtman 1998, p. 802). For example, a person with a good grade in math is at the top of his class and, by definition, belongs to the high-status group. However, if his class is ranked on sports activities, he may be in the lower status group if he is not good at sports. Or, if he meets other math enthusiasts at a competition, he may not be among the best because other math geniuses are better. Status[28] is thus a relational construct in which people compare themselves with others in their (peer) group. For example, people with A-level[29] differentiate themselves from people without A-level, but compare themselves with other A-levels, leading to different status positions.

As the definition by Ball et al. (2001) highlights, status must be recognized by society (or the group to which it belongs). Additionally, to be successful (or valid), some form of acceptance by the entity is required. When high- and low-status groups agree on the status structure, it is legitimate and can serve as a sanction or incentive mechanism (Weiss and Fershtman 1998, p. 802; Bettencourt et al. 2001, p. 521).[30] A socially recognized and accepted status leads to social validity, which is the basis for legitimacy. With a legitimate status structure, inequalities between different social groups are more likely to be accepted because they are justified (Ridgeway 2001, pp. 257-258, 273–274). As the legitimacy of status is important to this study, it is further elaborated in Section 4.2. The following definition specifies how status is used in our study.

[28] Status and social status are used synonymously in this study.
[29] In Germany the corresponding degree is the *Abitur*.
[30] See Mullen et al. (1992) or Bettencourt et al. (2001) for a meta-analysis concerning the legitimacy of status.

Definition 7: Status

The relative position of a person in his group, which is recognized and accepted by other group members, is his status, with high-status people being entitled to a favorable handling.

"One definition of social identity refers to a person's sense of self, derived from perceived membership in social groups. When we feel that we belong to a group, we may very well derive at least a portion of our sense of identity from that group" (Charness and Chen 2020, p. 692). Thus, *social identity* is a self-concept and guides individual behavior. The self results from different roles or positions in society or groups, so social identity and social status are closely related. According to social identity theory, originally developed by Tajfel and Turner (1979), three components of group differentiation are relevant. (1) In *Categorization*, people are classified by who they are associated with, like female or male. (2) *Identification* means association with these groups, that is, identification with one's group. (3) *Comparison* specifies that people compare their group, the in-group, to the other group, the out-group. In doing so, the in-group is seen as more favorable, which leads to the well-known in-group bias (Tajfel and Turner 1979, p. 41). Thus, membership in a group is meaningful for people and influences their behavior in accordance with existing group norms and values (among others, see e.g. Tajfel and Turner 1979; Shih et al. 1999).

Definition 8: Identity

By being member in a group, people derive their identity, thus their sense of self.

The *socioeconomic status (SES)* determines the social class of an individual compared to others. The social class defines the rank within a society and an individual's material resources. Thus, its implications are widespread and affect various dimensions, such as health and well-being. The literature distinguishes two forms of SES. The objective SES builds on factors such as educational opportunities or occupation. As this objective measure cannot account for the complex dimension of social class, researchers developed a more subjective measure. The subjective SES considers an individual's sense of place in society based on available resources (Kraus et al. 2009, pp. 992-993). Usually, the objective SES relates to differences between educational levels and not within one of these levels (in Germany, for example, the different school-leaving qualifications, like 'Realschulabschluss' or 'Abitur'). For the purpose of this study, it is logical to speak of status and not of SES, as the ranking

depends on the final grade of persons with Abitur and not on the different educational levels in general.

Definition 9: Socioeconomic Status

The SES means the social class that defines the rank within a society and individuals (material) resources.

In Table 4.1, the three hierarchy-related constructs, as well as their main characteristics, are summarized.

Table 4.1.: Hierarchy-related Constructs

Construct	Main characteristics	Based on
Status	ranking in a society, socially recognized, orients people outwardly	skills, accomplishments
Identity	person's sense of self	derived from perceived membership in social groups
SES	define the rank and material resources of the individual within a society	social position or class, based on either objective or subjective criteria

Implications of Group Rankings

If a group of people is differentiated, usually the part with the higher position has power, prestige, or can influence the others. *Influence* is a special dimension that often results from the previously mentioned hierarchy-related constructs. Influence is the ability to affect the behavior or views of others, either positively or negatively, without them necessarily being aware of where the influence is coming from. Even if individuals have influence, it does not automatically mean that they use it. Additionally, influence is not inevitably connected to a hierarchy, as the mere presence of another individual can already lead to influence (Blader and Chen 2014, pp. 74-75). Influencing a person always happens intentionally, which means that someone cannot influence without knowing it. This does not mean that the intentional act actually influences the behavior in the intended direction. For example, a default option that is already preselected (probably) influences the decision in the default direction but it

can also induce a completely different behavior. As French and Raven (1959) highlight, an influential action can also be the (passive) presence of a police officer at a corner, as this can already be perceived as an act of speed limit control (p. 152).

Influence is often closely associated with power, as someone with power can influence others. The main characteristic of *power* is that it is linked to control or coercion, which does not automatically imply a negative connotation. People with power (can) use pressure or threatening behavior to enforce their desires or get others to do what they want. However, they can also use their power without coercion, typically, to implement intended behavior by exerting control over others, either as sanctions or rewards (French and Raven 1959, pp. 152-153). An interesting aspect related to this study is legitimate power. According to French and Raven (1959), it is one base of power and further discussed in Section 4.2.2. Although status and power are related and both can exert influence, they are not the same and do not (automatically) occur together (Albrecht et al. 2013, p. 4). In this context, Hays and Blader (2017) summarize that status and power often covary in many social settings. However, status orients people outward because it is conferred through a social exchange process, whereas power orients people inward and diminishes their concerns about others (pp. 77-78).

Like power, *dominance* is associated with the use of coercion or threat that leads to a higher social rank by inducing fear in others, either through aggression, pressure, or withholding resources (Henrich and Gil-White 2001, p. 166; Blader and Chen 2014, p. 74). Another dimension that is closely related to status, or rather the result of a high status, is *prestige*. Being respected by others because of one's skills or competence can lead to a high rank and thereby prestige. If skills or competence help to achieve collective goals of the group, this is even more obvious (Blader and Chen 2014, p. 74). According to Henrich and Gil-White (2001), prestige builds on merit, implying that people with prestige deserve their superior position without inducing fear or exerting coercion (p. 170).

In economics, entitlement and property rights are prominent concepts and often used interchangeable, although they differ in some respects. Hoffman and Spitzer (1985) define *entitlement* as "legally enforceable claims (in a real-politic sense) to economic resources" (p. 260). However, this does not imply that an entitlement is in itself a legal right, even if it is (usually) accepted by others. What constitutes an entitlement is a subjective belief to earn and defend that right. The counterpart to entitlements are obligations (Schlicht 1998, p. 24). According to Lerner (1987), people often feel that the world is a fair place. Everyone (should) get what they deserve and a fair and legitimate social system (should) rewards effort. Thus, individuals or groups are entitled to an outcome based on their characteristics or previous

Table 4.2.: Implications of Group Differences

Implication	Main characteristics
Prestige	induces respect, based on merit
Power	coercion and/or control, threatening behavior, orients people inwardly
Influence	affect the behavior or the point of view of others
Dominance	fear or coercion
Entitlement	right or claim to economic resources, based on acquired characteristics in the past
Property right	morally justified entitlement
Decision right	right to decide

actions (p. 108). In distributive justice, entitlements are often identified as a justice criterion and defined as ascribed or acquired characteristics (such as social origin or sex), or status characteristics, which have been acquired in the past, such as occupational status. In this context, something is considered just if it is allocated based on these entitlements (Hülle et al. 2018, p. 668). It follows that people who are entitled to some right (may) deserve a reward or access to (more) resources. The conclusion that they receive more or take more money for themselves is logical and called the entitlement effect (Demiral and Mollerstrom 2018, p. 341). Section 4.3.1 provides experimental evidence of the entitlement effect.

According to Hoffman and Spitzer (1985), a *property right* is a "morally justified entitlement" (p. 260). Defined by the right, the holder obtains the right for a particular action (or set of actions) or is otherwise restricted in his acting. A property right makes the action of its holder fair and acceptable, insofar as it is legitimated by society and socially accepted (Hoffman et al. 1994, p. 350). The term decision right is often used in a broader context and generally means the right to decide. Bartling et al. (2014b) exhibit that many people have an intrinsic motivation to decide and thus value decision rights beyond any material benefit (p. 2005). Table 4.2 summarizes the definitions presented earlier.

4.1.2. Status and Identity in Psychology and Sociology

Status and identity have their roots in sociology and psychology and are addressed from the perspective of these disciplines. Of course, there are many different theoretical approaches to identity, status and hierarchy-related constructs. As it is impossible to present them all, two theories, Social Identity Theory (SIT) and Status Characteristics Theory (SCT), are discussed here. Both are relevant for this study and provide meaningful insights into how people are differentiated and the consequences of that differentiation.

Social Identity Theory

A prominent line of research in sociology and psychology is concerned with the social identity of individuals and its effects on behavior toward people with the same (different) identity. In the previous section, a definition of social identity was given. In summary, belonging to a group matters to people and determines their social identity. Based on norms, their social identity defines the appropriate behavior for different social groups, so people do not deviate from these group norms. They exhibit an in-group bias insofar as their own group receives preferential treatment (Paetzel and Sausgruber 2018, p. 281).

The SIT also explains how people react to social status. The status (structure) plays an essential role in defining the behavior between groups, with significant differences in how people respond to those in the in-group or out-group. Status is considered to be the result of comparisons between different groups, where individuals try to achieve a higher status, thereby verifying their identity. In this regard, comparisons with high-status groups are more favorable as they reinforce a positive social identity. In contrast, lower status and comparison with low-status groups negatively affects social identity (Tajfel and Turner 1979, p. 43).

Akerlof and Kranton (2000) incorporate the concept of social identity into economics. By implementing it into a utility function, they show how economic outcomes are affected by (social) identity. Based on their work, Shayo (2009) proposes a theoretical model of social identity, where two components or processes determine identification with the group. The first is social status, which is the relative position of the group on various dimensions of comparison, such as occupation or income, while identification is more accessible with high-status groups. The second one is the (perceived) social distance between oneself and the other group members. The classification as a group member is more straightforward as more similarities to group members exist (p. 147). Shayo (2009) specifies that "identification with a group means caring about the status of that group" (p. 151).

Numerous experimental studies show that even a trivial difference between groups is sufficient to prefer people of the same group, confirming the in-group bias. Whenever a group perceives itself as different from others, it values their in-group as better than the out-group and act according to their group, even when the group differences are meaningless (among others, see e.g. Tajfel and Turner 1979; Ridgeway et al. 1998; Paetzel and Sausgruber 2018; for a review, see Hewstone et al. 2002). Further research highlights that belonging to a group has value to people and induces psychological benefits. For example, people would pay money to stay in their preferred group (Heap and Zizzo 2009). Additionally, the in-group bias is stronger for higher status people than for low-status people, especially when the dimension

of comparison is favorable to one's own group (Brewer and Kramer 1986).

Butler (2014) compares the effect of a status distinction among different identities and the implications on behavior in a trust and cheap talk game. Participants are randomly assigned to two distinct color groups (= identity), one of which is (randomly) declared to be of higher status. In a baseline treatment, the colors are still present, but a declaration of status differences is left out. As a result, he exhibits that status differences reduce the in-group bias. Consistent with the theoretical predictions of Shayo (2009), Hett et al. (2020) emphasize that groups with lower social distance and higher social status are preferred. Paetzel and Sausgruber (2018) find that the in-group bias in minimal groups[31] is high and significant for participants with relatively low status. Consequently, high-status groups show less in-group bias than low-status groups. In a meta-analysis, Mullen et al. (1992) find out that the in-group bias for high-status people exists only when social groups are artificially created, that is, by the experimenter. When status differences are real, the in-group bias of high-status groups disappears. Apart from these studies, there is a vast body of research addressing the effects of social status on identity.[32]

Status Characteristics Theory

Another interesting and noteworthy psychological theory is the SCT, originally developed by Berger et al. (1966, 1972). The SCT is concerned with the influence of status characteristics on the organization of social interactions. Status differences are based on individual characteristics and determine who acquires power or prestige in a group task (Berger et al. 1966, p. 5, 1972, p. 241).

In their sense, a status characteristic can be any characteristic, such as age, race, sex, skills, or abilities, which has at least two different states. The theory therefore assumes that these states are evaluated differently in terms of desirability, honor, or esteem and are associated with varying performance expectations. Even if the characteristics are not directly relevant to the performance in a particular task, they have a strong effect on the

[31] In laboratory experiments participants are often assigned to (artificial) minimal groups in order to avoid systematic differences and implement randomly determined groups (among others, see e.g. Akerlof and Kranton 2000; Heap and Zizzo 2009; Paetzel and Sausgruber 2018). To do so, the Minimal Group Paradigm based on the work of Tajfel et al. (1971) is often used, which aims to divide participants as randomly as possible into different groups. Therefore, typically, the preferences for two painters, Klee and Kandinsky, whose paintings are quite similar, is used. In the original experiment, participants have to decide which one they prefer among 12 pairs of paintings, without knowing which painting belongs to Klee or Kandinsky. Afterwards, participants are randomly assigned to the Klee or Kandinsky group (p. 165).

[32] Concerning identity in economics, a good overview of recent approaches, results and related issues is provided by Charness and Chen (2020).

distribution of power in group hierarchies (Berger et al. 1966, p. 5). Performance expectations are based on assigned status characteristics. For example, individuals ascribed higher ability are expected to perform better than those with lower ability (Berger et al. 1972, p. 246). Berger et al. (1966) summarize it as follows: "a status characteristics may be thought of as any characteristic that has differentiable evaluated states that are associated directly or indirectly with expectation states" (p. 35).

Berger et al. (1966, 1972) distinguish two forms of status characteristics: diffuse and specific. A diffuse status characteristic, such as race or gender, is not directly linked to expectations on a specific task.[33] A specific characteristic, on the other hand, encompasses only a set of expectations relevant to a particular type of task, such as math abilities in a math task. Specific characteristics are likely to be related to a person's productivity and thus convey information to others, influencing expectations and behavior toward them.

According to the SCT, the meaning and value of status characteristics are attributed by society and arise in social interactions with others. As with status, the effects or implications of status characteristics depend on the situational context. In order to be effective, they need to be activated. Berger et al. (1972), however, do not define the conditions necessary to activate them (p. 244). Applying the SCT to this study, the final grade is a specific rather than a diffuse characteristic. It could be argued that a better grade is linked to more effort and (or) higher cognitive ability, which can indicate more foresight and thus the expectation of a better or more equitable allocation of the endowment.

Although the SCT has not been free of criticism, several empirical studies confirm the effects of status characteristics on performance expectations and the corresponding behavior of individuals (see e.g. Ridgeway et al. 1998; Kalkhoff and Barnum 2000; Simpson and Walker 2002; Hong and Bohnet 2007).

Summary 4.1: Status in Psychology

To summarize, the SCT deals with status characteristics, which means a characteristic consisting of two or more states for which a society (or a group) shares the same valuation of these states. Therefore, it is socially desirable to achieve one state over the other(s). In contrast, group membership in the SIT is not based on shared values. Even a trivial (random) affiliation to a group already leads to the in-group bias. Consequently, the SCT and the SIT deal with different ideas of social influence.

[33] Yet, it could be argued that women are generally considered to be better at math, so there might be an expectation that women will do better on a math quiz.

However, it is far from self-evident that they do not interact or influence each other (Kalkhoff and Barnum 2000, pp. 95-96). Kalkhoff and Barnum (2000) examine in a laboratory setting whether the two theories overlap. Overall, they conclude that both concepts determine social influence to the same degree. Regardless of group membership, a person with higher social status has more influence than someone of lower status because of the expectation of higher ability. Additionally, regardless of status, members of the in-group have more influence than members of the out-group because of their similarity (pp. 110-113).

4.1.3. Status in Economics

Recently, there have been several attempts to implement (social) status into economics as it is evident that social considerations have implications for economic decisions. For example, the status of group members can influence the distribution of resources affecting the group's efficiency (see. e.g. Fershtman et al. 2012).

Social status in economics is mainly characterized by two dimensions. First, as defined earlier, social status is positional and leads to direct consequences in the form of negative externalities after consumption. Consequently, people often prefer to obtain a high(er) status. When one part of an entity (a person or a group) achieves a higher status, the other part of that entity must decrease in status. Thus, as in a typical zero-sum game, the (status) gain of one part is the loss of the other part (Weiss and Fershtman 1998, p. 802; Heffetz and Frank 2011, p. 73). Second, status is an intermediate good that is part of the individual utility function. Hence, it can have the same value as a non-monetary currency that gives a person with a certain status access to resources and higher utility. These benefits (in terms of utility) that arise from status differences need to be considered (Ball et al. 2001, pp. 162-164). Thus, status has a value in its own. Having a higher status is equivalent to reaching utility as it is transferable to resources bringing utility.

The literature consistently emphasizes that people want their outcomes to be proportional to social status. The acceptability of lower incomes for low-status workers compared to higher incomes for high-status workers is relatively high, contradicting equality considerations (Kelley and Evans 1993, p. 115). This is further supported by the results of the European Social Survey from 2018[34], where people get information of the highest and lowest incomes in their country and have to assess their fairness on a 9-point scale. People with a higher income are also more likely to rate their (high) income as fair. In contrast to the other

[34]The ESS has been conducted since 2002, while justice and fairness criteria of income distributions were collected in 2018 (Adriaans et al. 2019, p. 404).

European countries, half of low-income workers in Germany (the lowest 20%) evaluate their income as fair. This positive attitude and higher acceptance towards one's own (low) income is probably driven by a high share of part-time worker in Germany (Adriaans et al. 2019). Based on the FS-model, Tutic and Liebe (2009) propose a utility function incorporating status-mediated inequality aversion, where outcomes should be distributed proportional to social status, effort, or ability. Thus, the fair reference point depends on the status of others, as this is the normative orientation for valuation (pp. 159-160).

However, even if people prefer to achieve high status, status is neither purchasable nor tradable, as it is assigned by society, the social system, or the entity to which one belongs (Heffetz and Frank 2011, pp. 73-74). Depending on how status is achieved, it can establish an entitlement to specific resources or privileges that lead to a favorable reaction from others. Consequently, someone with higher status has more authority, is the leader, or is generally seen as more deserving, which is accepted by others with whom he interacts. As a result, there is often a positive correlation with a person's social status and his income (Weiss and Fershtman 1998, p. 802; Ball et al. 2001, pp. 161-162).[35] The mere distinction or classification into different status positions gives one group an advantage over the other as beliefs are formed. Usually, one category or position is ascribed more competence or social worth than the other. Furthermore, a (social) status can provide (economically relevant) information (Ridgeway 2001, pp. 257-258, 273). In his popular theory of conspicuous consumption, Veblen (1899) highlights that status has an intrinsic value as a resource in itself and can be expressed by the consumption of precious goods (among others, see e.g. Heffetz 2011; for a theoretical analysis see Bagwell and Bernheim 1996).

The characteristics of status presented so far already imply one of the main benefits of a status position: a high status can lead to a higher income, which creates the desire for a higher status. Status-seeking behavior is rational behavior when the expected benefits of a better status position are higher than the corresponding costs to achieve it (Ball and Eckel 1998, p. 502). However, even in the absence of monetary rewards or benefits, the mere existence of higher status can induce an intrinsic value to a person. Consequently, the (social) rank is relevant to individuals even without direct (measurable) economic benefits (Weiss and Fershtman 1998, p. 802).

[35] Of course, this relationship is twofold. Income also correlates with individual characteristics that imply high status, such as education, intelligence or the origin.

Summary 4.2: Status in Economics

Status in an economic sense is a valuable good that is assigned by the entity an individual belongs to and thus, neither purchasable nor tradable. As it is part of the individual utility function with a higher status leading to more access to resources, people seek to attain a higher status increasing their individual utility. A higher status provides information about the status holding person as he (may) have the favorable position due to more competence or ability.

Based on these theoretical considerations on status, we use the following working hypothesis to specify implications and expectations for our experiment.

Working Hypothesis 11 (High-Status) *People with a higher status may be perceived as more competent to distribute resources in their group.*

4.2. Theoretical Basis of Legitimacy

As a moral concept, (normative) legitimacy is a core element of social interactions between people, groups, organizations, or societies. The word legitimacy derives from the Latin word 'legis', which means 'according to the rule' (Zelditch 2001, p. 33). Consequently, it is associated with the belief that an actor, e.g., an authority or institution, is just, so that people voluntarily accept it and its actions. Whenever people see an authority as legitimate, its decisions are evaluated to be right, and a personal feeling of obeying results (Tyler 2001, p. 416).

4.2.1. Normative and Empirical Legitimacy

The literature focuses on legitimacy as a normative concept that indicates what should be legitimate and what standards are considered desirable. However, to fully understand the concept of legitimacy and its implications, the normative view must be complemented by an empirical perspective that captures how people perceive legitimacy. A popular negative example of how empirical and normative legitimacy can lead to different approaches is the Nazi regime. Although they had the support of the people (empirical legitimacy), they did not reach the basic level of normative legitimacy (Hough et al. 2013, pp. 326-330). This section addresses and distinguishes between these two concepts of legitimacy.

Legitimacy as a *normative concept* involves objective criteria. In order for something to be considered legitimate in the normative sense, predetermined standards must be achieved. Therefore, it is not relevant whether people consider these standards to be satisfied or not. When people speak of normative legitimacy, they do not mean that they are subjectively convinced of the legitimacy of something, but rather that the essential requirements have been met (Hinsch 2008, p. 41).

Definition 10: Normative Legitimacy

Normative legitimacy means having achieved predefined standards.

Hinsch (2008) concludes that "to say that an institutional arrangement is legitimate in the normative sense is to publicly recognize that it has moral standing and not simply to report that there are people who believe that the arrangement has moral standing" (p. 41).

The central aspect of *empirical legitimacy*, which goes back to Max Weber, is that norms or regulations meet with the approval of the people who have to live with them. Consequently, they support them voluntarily and evaluate them as right because they are relevant. However, this does not mean that they are justified in a normative way. As the example at the beginning highlights, an authority can be seen as legitimate in an empirical sense and at the same time be rejected for other (normative) reasons (Hinsch 2008, p. 40).

Definition 11: Empirical Legitimacy

Empirical legitimacy means having the approval of the people.

Therefore, in an empirical sense, "to say that an institutional arrangement is legitimate [...] is to make a factual claim about the subjective state of mind of particular individuals that belong to one political society" (Hinsch 2008, p. 40).

4.2.2. Roots of Legitimacy

Legitimacy has a long history, especially in political sciences. Aristotle, Machiavelli and other classical philosophers, highlight the importance of legitimacy in the context of political stability.[36] For the stability of a political institution, the (voluntary) acceptance of it is a prerequisite. To achieve political acceptance and, thus, political stability, the (empirical)

[36]Zelditch (2001) provides a chronological overview of theories of legitimacy (pp. 35-36).

legitimacy of the institution is fundamental (Zelditch 2001, pp. 36-37). As Tyler (2006) points out, "effective democratic governance depends upon the legitimacy of the state" (p. 380). In this regard, legitimacy is a form of power that allows authorities to influence people's behavior without using rewards or sanctions. Legitimate authorities, institutions, or governments benefit from having legitimacy in the form of less effort and more efficiency (Tyler 2006, pp. 376-377).

According to French and Raven (1959), the basis of legitimacy is internalized norms or values that establish that the actor is legitimate to influence or control the people. Consequently, legitimacy involves a certain right or standard accepted by others. Of course, this is not necessarily to be understood negatively, as an influential act can also lead to positive results. Additionally, legitimacy is a base of power. In this regard, the internalized values induce the actor to have power because of the legitimate right to influence others, who must accept this in return (p. 159). Three components form the basis of legitimate power: (1) cultural values, (2) acceptance of the social structure, and (3) designation by a legitimate actor. The second component, in particular, is relevant in this study. When people accept the status structure, they also accept the legitimate authority (here, the decision-maker) and thus his decisions (French and Raven 1959, p. 160), which is discussed in more detail at the end of this section. Another conceptual distinction of legitimacy considers different levels of acceptance: the collective and the individual level. For example, at the collective level, a policy may be considered legitimate because it is seen as valid by some people (but not all), whereas at the individual level, not everyone agrees (Tost 2011, p. 689).

As for the concept of causality, legitimacy is typically evaluated on two dimensions: (1) the legitimacy of the action and (2) the legitimacy of the actor, i.e., the one performing the action (Kelman 2001, p. 55). Assuming the actor uses fair procedures and is seen as legitimate, unpleasant outcomes are more likely to be attributed to external circumstances rather than to the actor himself. Additionally, if the actor is seen as legitimate, meaning that he has the support of others (= empirical legitimacy), unpleasant outcomes are more likely to be accepted as they are perceived as less unjust (Hegtvedt et al. 2003, pp. 355-356). To evaluate whether something is fair orients a person outwardly as the evaluation is (usually) based on the behavior of others. People rely on the justice and legitimacy of the procedure when making attributions for decisions and outcomes (Hegtvedt et al. 2003, p. 344). Consequently, the voluntary acceptance that results from legitimate procedures is twofold. The executed action and the status structure are more likely to be accepted if they are perceived as legitimate. In the context of this study, it implies that a decision-maker has the right to decide and that the affected party (the recipient) recognizes this legitimate right.

More specifically, the affected accepts something as right, which leads to stable structures.

However, legitimacy can also lead to negative effects, as a legitimate authority can abuse or oppress others. The power that authorities gain through a legitimate process can give them the influence to control the behavior of others and act in immoral ways (Tyler 2006, p. 382).

Summary 4.3: Legitimacy

A decision-maker can be legitimate in an empirical sense if the others agree with his (outstanding) position and support him as the appropriate decision-maker. However, to be legitimate in a normative sense, objective criteria have to be defined and met, which does not mean that all people consider these criteria to be met.

The definition of the final grade as an objective criterion, according to which the decision-maker is declared to be the person with the decision right, is fulfilled in this setting, so normative legitimacy is achieved. The final grade is legitimate in a normative sense as it follows predefined standards by the Ministry of Education. Therefore, it is not questionable if the final grade is legitimate per se or not. Rather, it is of main interest how the role-assignment in a laboratory experiment, which is chosen by the experimenter, is perceived as legitimate or not. In laboratory experiments there are no predefined standards by an institution or a moral code of how roles or positions are determined in a legitimate way. Thus, the evaluation of the final grade as legitimate in an empirical sense, is the main point of our experimental design. For example, if participants do not agree on the final grade as a criterion, the decision-maker is not necessarily legitimate in an empirical sense. Alternatively, in the random treatments no objective criterion is met. People might agree on the decision-maker's position in an empirical sense because he is declared to be the decision-maker and they support him, whereas normative legitimacy is not necessarily given.

Further evidence on experimentally tested legitimacy of status differences is given in the next section.

4.2.3. Legitimacy of Status

Legitimization is not restricted to institutions or authorities. Instead, many social processes at different levels are legitimate, such as status hierarchies, inequalities, wealth, or rewards (Zelditch 2001, p. 34). Thus, the idea of legitimacy encompasses many social settings and determines the position and organization of social entities. It is part of many social processes

that assign status or power to individuals, groups, or societies. When there are social or economic differences, questions are almost always raised about their legitimacy and the process by which these differences arise (Tyler 2006, p. 385).

A vast body of the literature examines the effects of more or less legitimate high- and low-status positions. For example, status differences are more likely to be accepted when they are perceived as legitimate, especially by low-status groups. Conversely, inequalities are more likely to be accepted when status differences are legitimate (among others, see e.g. Olson and Hafer 2001; Levin et al. 2002). Consequently, the acceptance of hierarchy differences depends strongly on perceptions of legitimacy. Low-status members identify more strongly with their lower status when a legitimate procedure implements it. In contrast, illegitimate procedures lead to more competitive behavior by low-status groups because they do not perceive high-status groups as more deserving than themselves (Ellemers et al. 1993, p. 767). When high-status people perceive their superior position as illegitimate, they may feel guilty, while low-status people may feel angry (Weber et al. 2002). Apart from the evaluation by the affected party, legitimacy is essential for one's own evaluation. Legitimate actions justify one's own behavior, feelings, and thoughts. As a result, a positive self-image is maintained even if an unpleasant allocation is chosen (Jost and Major 2001).

As it is already highlighted, different categories, such as occupation, education, or gender, are associated with different personality traits and abilities, and form the basis for expectations and norms. Decisions that conform to these expectations or norms are thus legitimate, since legitimacy builds on these norms (Zelditch 2001, p. 49). People with higher (legitimate) status may have favorable characteristics and attributes, and are seen as more competent and smarter (Tyler 2001, p. 416). A legitimate group status confers superiority and makes them more deserving than others. Consequently, if a superior group (or group-member) decides to implement a less profitable outcome for the inferior group, it is more likely to be accepted because the superior group deserves the better outcome (Turner and Brown 1978).

Summary 4.4: Legitimacy of Status

The perceived legitimacy of a group status determines the reactions toward and from other group members. Legitimate social systems are more likely to be accepted, even if they lead to inequalities. Thus, the acceptance of an unequal distribution of resources depends strongly on the legitimacy of the underlying system, which has to be seen as legitimate by both sides – by those in an advantaged and those in a disadvantaged position. Only then are the advantages of a higher status position beneficial and usable.

4.3. Status Differences in Laboratory Experiments

In many laboratory experiments, subjects are randomly assigned to treatments and roles so that a(n) (un)favorable position is usually obtained by pure luck. This role-assignment procedure can be interpreted as unearned, undeserved, or illegitimate. Consequently, participants with a higher position may be reluctant to use their randomly assigned power, while low-status people are unwilling to accept the decisions of the high-status group (Feltovich 2019, p. 153).

The following sections are devoted to exploring two things: first, relevant experimental results examining status differences and role assignment procedures are presented. Secondly, an overview of typical group-building mechanisms in laboratory experiments is provided.

4.3.1. Experimental Evidence on Status Differences

In experimental economics, the entitlement effect is quite common. One of the first papers analyzing it is by Hoffman and Spitzer (1985). In their experiment, the winner of a coin flip decides whether he wants to have 12 dollars for himself, giving nothing to the other participant, and alternatively, whether he wants to cooperate with the loser so that he receives 14 dollars and can propose a distribution of the endowment. To be paid out, this proposed distribution must be accepted by the looser. Contrary to the game theoretic predictions, all winners decide to cooperate and divide the amount equally between themselves and the other participant. Hoffman and Spitzer (1985) conclude that the assignment of the decision (either as winner or loser) affects self-and other-regarding preferences. Consequently, random role assignment through a coin flip might be perceived as unfair or illegitimate, so the winner does not see himself as entitled to take more.

The seminal paper of Hoffman et al. (1994) extends this work and highlights that an earned decision right has vast implications on distributive decisions in dictator and ultimatum games. Overall, the decision-makers behave more selfishly when they earn the right to decide in a general knowledge quiz than when they are randomly assigned (pairwise Epps-Singleton-Test, $\chi^2 = 10.4, p = 0.03$). Although offers in the ultimatum game differ between random and earned decision rights, there are no significant differences in recipients' rejection rates, with only 2 of 24 offers rejected in randomly assigned groups and 0 of 24 rejected with earned decision rights (pp. 12-17). In their design, the role-assignment procedures do have an effect, but only on the dictators and not on the recipients.

Originally, Demiral and Mollerstrom (2018) intended to examine gender differences in the ultimatum game using the experimental design of Hoffman et al. (1994). Interestingly, al-

though their design is similar, they cannot replicate the main results of Hoffman et al. (1994). They find no significant differences in proposer behavior as a function of role assignment, with randomly assigned proposers offering an average of 41% of their endowment, while proposers with earned roles share 42% of their endowment. A two-tailed t-test assuming unequal variances reveals no statistical differences ($p = 0.182$). Furthermore, they analyze recipients' rejection rates, finding no significant differences between both procedures (both groups state a minimum acceptable offer of 28% of the endowment, t-test, $p = 0.706$) (pp. 346-348). Although their robustness checks, such as assessing the deservingness of proposers and recipients, show that proposers are rated as more deserving (scale from 1 to 10, 7.23 in the earned role treatments vs. 6.52 in the randomly assigned treatments, t-test, $p < 0.01$), no differences occur that stem from different group-building mechanisms (Demiral and Mollerstrom 2018, p. 346).

Another related experiment focusing on status differences in the ultimatum game comes from Ball et al. (1998). Participants answer questions in an economic trivia quiz and are divided into high- and low-status groups based on the number of answers (rather than correctness). They use a 2x2 design in which the proposer is either of high- or low-status and is matched with a recipient from the other status group. Then, the proposer decides how to divide either \$10 or 10 Hershey's Kisses. A control treatment in which status positions are randomly assigned is also implemented. High-status individuals are considered more deserving and are treated better, regardless of their position as proposers or recipients. Low-status proposers are more generous towards high-status recipients and vice versa (Epps-Singlton-Test, significant at the 5%-level). Interestingly, this superior handling of high-status people is also evident in the random treatment, contradicting the findings of Hoffman and Spitzer (1985), but aligning with Demiral and Mollerstrom (2018).

Further experimental results highlight that the status position influences the acceptance of income inequalities, in that low-status people are more satisfied with disadvantageous payoff differences than high-status people are (Albrecht et al. 2013). Relating to this, Fleiß (2015) examines the impact of merit in an ultimatum game. Roles are either randomly assigned (participants draw cards with numbers, which they assign to computers with predefined roles) or acquired in a general knowledge quiz, with the winner declared as proposer. In summary, their results are consistent with previously presented experiments, in which proposers offer less when roles are earned (random: 7.6 vs. earned: 6.5, U-test, two-tailed, $p = 0.040$), while recipients accept less in the earned treatment (4.7) than in the random treatment (6.5), which is significantly different at the 1%-level (U-test, two-tailed, $p = 0.009$) (pp. 397-400). Although overall rejection rates do not differ between role-assignment procedures

(U-test, two-tailed, $p = 0.934$), rejection rates for offers below the 50-50 split are significantly lower when participants earn their roles (0.313) than when they are randomly assigned (0.506), which is statistically significant at the 1%-level (U-test, two-tailed, $p = 0.005$). Hence, recipients accept to receive less when roles are earned because they (may) view others with better performance or higher status as more legitimate and are therefore more willing to accept their decisions (Fleiß 2015, pp. 389-390). Similarly, using an ultimatum game with status differences, Blue et al. (2016) confirm these results with both one's status and the status of the interacting person influencing acceptance behavior. They highlight that recipients of lower status are more likely to accept lower offers and that low offers are more likely to be accepted by high-status decision-makers.

Huberman et al. (2004) examine the role of attaining higher status in different cultures and exhibit that status is valued independently of monetary aspects. People are willing to pay for a higher status, while this status-seeking behavior is more pronounced for men than for women. In a dictator game with third-party punishment, von Essen and Ranehill (2011) induce status by a noble or common surname showing that low-status men with common names are punished more than low-status women. Other experimental and empirical studies support the connection between gender and status-seeking behavior, showing that men are generally more sensitive to social hierarchies (among others, see e.g. Campbell 2002).

Even in moral disengagement situations, the interacting person's status influences violent behavior, here receiving money for inflicting electric shocks. Among participants matched with someone of the same status position, i.e., high status matched with high status or low status matched with low status, fewer electric shocks are carried out than when matched with someone of a different status position (29.6% with equal status and 74.1% with different status), which is statistically significant at the 1%-level (Falk 2017). In bargaining situations, Ball et al. (2001), Feltovich (2019), and Gächter and Riedl (2005) point out that an earned bargaining position results in less equal splits and is, thus, closer to the theoretical predictions than with randomly assigned roles.

Summary 4.5: Experimental Evidence on Status Differences

To summarize, these results indicate that one's status and the status of the interacting person influences the acceptance of resource distributions, with participants in earned roles generally deserving more. Participants are more willing to accept income inequalities when status differences exist. However, the way roles or positions are assigned, either randomly or through a specific procedure, does not automatically lead to significant differences in allocation decisions and reciprocal reactions. Three interesting aspects are worth noting:

(1) When high- and low-status individuals interact, high-status individuals tend to be treated better, are seen as more deserving, and may allocate a smaller amount to recipients.

(2) This superior handling of high-status people is voluntarily accepted by low-status people.

(3) There may be an effect of the role assignment procedure on reciprocal reactions that is prevalent only for low offers or amounts under the equal split.

Consequently, we expect people with lower status to accept a smaller amount for themselves, as people with higher status deserve more. Since allocations in our setting are either equal (everyone gets the same) or unequal (dictators get almost all of the endowment), the process by which roles are determined is expected to influence the response of recipients. Because unequal allocation distributes only a small fraction of the resources to recipients, differences in reciprocal responses between a random and a nonrandom procedure are likely to occur. Together with the evidence presented on legitimate status differences, the following working hypothesis states the resulting expectations for our experiment.

Working Hypothesis 12 (Role-Assignment) *We assume that legitimate status differences lead to a higher acceptance of inequalities and consequently fewer punishment points are imposed for decision-makers with a legitimate status position.*

4.3.2. Implementation of Status Differences

As it is already specified in the last section, there are various procedures or mechanisms by which individuals are assigned to groups, roles, or positions that induce status differences in laboratory experiments. They range from more or less random mechanisms to real-effort-

tasks, which require participants to work in some way, over to different quizzes, either as knowledge, trivia quiz, or mathematical tasks.

Table 4.3 overviews different group-building mechanisms typically used in experimental settings. It should be mentioned that this overview is far from comprehensive and classifies tasks based on their similarity, although other classifications are also conceivable.[37] Five categories of tasks are distinguishable:

(1) Random procedures, in which group-building is as random as possible, for example, by a coin flip or a computer.

(2) Real-effort tasks that require effort. One prominent example is the slider task or the hash-mark game. These tasks are often used when participants have to earn their endowment during the experiment.[38]

(3) Real-effort tasks of cognitive ability such as Raven's Matrices or math tasks.

(4) Real-effort tasks of quizzes, such as knowledge or trivia quizzes. The last two categories are closely related as both require some knowledge or skill.

(5) Characteristics acquired outside the laboratory, such as age or gender.

In laboratory experiments, there are two ways to have participants solve a task. In one, they are divided into pairs and compete against each other. After completing the task, one of them is declared the winner and is given a superior role (for example, as in Hoffman and Spitzer 1985). Alternatively, all participants compete in one session and are then ranked according to their performance (for example, as in Demiral and Mollerstrom 2018). If the task is completed simultaneously and participants are ranked according to their performance, then the conditions of the status definition from the beginning of this chapter are satisfied; a ranking of people, which is socially recognized and associated with a superior handling or reward, here in the sense of the decision right.

To indicate different social ranks and, thus, high- or low-status positions in the laboratory, different methods are usually used. In one, stars are assigned to participants (among others,

[37] This overview focuses on experiments from economics; there are, of course, more procedures from other disciplines such as psychology, often involving deception.

[38] This has vast implications for subsequent decisions – for example on the (re)distribution of endowments or charitable giving. For example, according to Dankova and Servatka (2015) and Cox and Hall (2010) the intensity of reciprocal responses is stronger when the endowment is earned. For an overview of this literature and corresponding procedures, see the article by Carpenter and Huet-Vaughn (2019). Since the endowment is not earned in this study, but distributed by the experimenter, we do not discuss these experiments further.

Table 4.3.: Overview of Tasks used in Laboratory Experiments

Category	Task	Exemplary paper
Random	Computer / Experimenter	Bartling et al. (2015), Anselm et al. (2022)
	Coin flip	Hoffman and Spitzer (1985)
	Color chips	Butler (2014)
	Klee-Kandinsky	Paetzel and Sausgruber (2018)
Effort	Slider	Gill and Prowse (2019)
	Hash-mark game	Hoffman and Spitzer (1985)
	Encoding words / coding speed test	Feltovich (2019), Segal (2012)
Cognitive ability	Raven's progressive matrices	Paetzel and Sausgruber (2018)
	Math/arithmetic task	Blue (2016), Demiral and Mollerstrom (2018)
	Summation task	Hett et al. (2020)
Quiz	Current event quiz	Hoffman et al. (1994)
	General knowledge quiz	Fleiß (2015)
	Economic trivia quiz	Ball et al. (1998; 2001)
Characteristics	Gender	Eckel and Grossman (2001), Campbell (2002)
	Name	Essen and Ranehill (2011)
	Age	Mertins et al. (2013)

see e.g. Ball et al. 2001; Blue et al. 2016) or participants are declared as member of a high-status (low-status) group (among others, see e.g. Paetzel and Sausgruber 2018; Hett et al. 2020). Alternatively, the status position is not explicitly declared, but indirectly implemented through an advantageous handling, such as being the decision-maker (among others, see e.g. Fleiß 2015; Demiral and Mollerstrom 2018). When a quiz assigns people to different groups or roles, participants are often referred to as winners or losers (among others, see e.g. Hoffman and Spitzer 1985). As these examples highlight, there are two ways to indicate status differences. Either participants explicitly know that they have high status and a superior role, or they have a special role but are not explicitly told.

In our experimental design, participants are not explicitly declared as having a higher status. Rather, the instructions explain that participants with a higher grade are decision-makers and have the right to determine the final allocation. To declare explicitly the superior role of participants might prime them and, as a result, their behavior might be influenced. Additionally, the declaration can foster the identification with the in-group, which is expected to result in more in-group favoritism and discrimination against the out-group. This would, probably, lead to more punishment by the recipients as they generally discriminate against

the dictators as their out-group.[39] As the aim of our study is not to further investigate the in-group bias and the influence of social identity, we decide to induce status differences indirectly. With this, we expect to measure the effect of status differences on the assigned responsibility and not identity implications.

[39] This would also imply that dictators generally decide in favor of their in-group, which would result in far more choices of the unequal allocation.

5. Upshot: How Should (Could) one Attribute Responsibility to Individuals in Groups?

In the introduction we highlight the main aspect of our study: how recipients assign responsibility to the decision-makers. More specifically, the recipients' reaction in terms of assigned punishment is determined not only by the decision making process, but also by the group formation process. Therefore, Part I highlighted the theoretical and empirical evidence on factors influencing this reaction, which we aim to bring together in this chapter in order to provide a synthesis of what determines assigned responsibility in a setting with (legitimate) status differences and actively (passively) implemented decisions.

The decision-makers in our experiment decide individually which allocation to choose, while the final outcome is the result of a group decision consisting of several individual choices. Consequently, we have two dimensions on which responsibility can be assigned: first, to the individual decision-maker for his or her choice, and (or) second, to the group as a whole, since no individual alone can determine the final outcome. The judgment as to whether a decision-maker is responsible for a choice and (or) the outcome, then provokes punishment.

Thus, in the simplest way, the following relationship can be assumed: a recipient compares the actually chosen allocation (or outcome) with (his) just perception, evaluates the decision-maker as responsible for a deviation, and thus punishes him. However, it is not so simple, and the relationship is not (always) as straightforward as described. There are factors that mitigate judgment and/or reaction, as a result of which someone is not punished even though the actual and fair allocation may differ. For example, if a recipient perceives the status differences as legitimate, he may be more likely to accept the responsibility of the decision-maker and assign no (or less) punishment. Consequently, even if a decision-maker is perceived as responsible, punishment may not be inevitable.

In what follows, we focus on the comparison between actual allocation and fair allocation, where the latter is guided by situational factors and individual preferences.[40] The result of

[40]For simplicity, we refer to decisions and outcomes using the term allocation, although we still have in mind the two possible dimensions (individual- and group-level) in which responsibility can be attributed.

the comparison is the judgment of whether a decision-maker is responsible or not, leading to the (potential) punishment reaction.

In our experiment, we can only measure the reaction (in terms of assigned punishment points), not the evaluation per se. Thus, we have to assume that participants punish when they judge someone to be responsible. However, as the Shultz-Schleifer Model (see Section 2.1.2) makes it clear, there is in fact a difference between assigning responsibility and punishing, which is not necessarily the same thing. However, without making this assumption, perceiving someone as responsible without punishing is not measurable with our experimental design, although of course it may be the case.

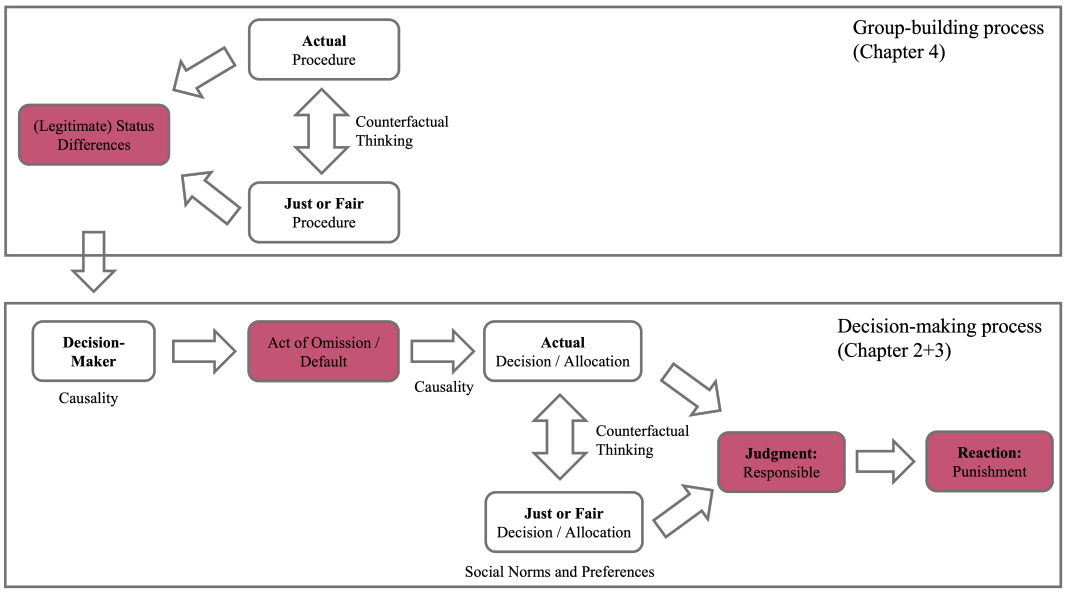


Figure 5.1.: Schematic Visualization of Responsibility Attribution

First, and as shown in Figure 5.1,[41], we distinguish between the decision-making process and the group-building process, and begin by focusing on the former. The comparison between the actual allocation and the just or fair allocation (which leads to the judgment that someone is responsible) is guided by social norms, preferences, and counterfactual thinking.

Related to this, (social) norms define the appropriate behavior in specific situations and thus change from situation to situation. Even small changes in a situation can alter the appropriate norms (Bernhard et al. 2006; Fershtman et al. 2012, pp. 137-140), thereby influencing individual behavior in almost all decisions. The appropriate norms in a given

[41] The visualization follows Jasso et al. (Jasso et al. 2016, pp. 201-203), who use a similar concept to specify how people form perceptions of distributive justice.

situation and others' expectations of the decision-maker acting in accordance with these norms then influence the punishment decision (Elster 1989; Fehr and Gächter 2000b; Ostrom 2000).

Counterfactual thinking is the mental consideration of possible (hypothetical) alternatives in order to evaluate whether a behavior is appropriate (i.e., in accordance with social norms) or whether some other behavior (i.e., a different allocation) would have been better, in other words (more) just (Kahneman and Miller 1986; Gurdal et al. 2013). Thus, counterfactual thinking is relevant in determining the cause of an outcome (Mandel and Lehman 1996, p. 450). The concept of causality, as defined by Orcutt (1952) or Hausman (1986), implies that manipulating the cause would lead to less (more) attributed responsibility and thus less (more) attributed punishment. This also implies that the relationship is unidirectional, since, for example, norms influence the decision to punish, but not the other way around.

Consequently, norms define the just allocation depending on the situation, while a recipient compares the actual allocation with the just allocation by counterfactual thinking of possible alternatives that determine the attributed responsibility. However, the concept of responsibility attribution, or rather the evaluation of whether a decision-maker is responsible or not, is a multidimensional construct, since different factors, sometimes exclusive, sometimes overlapping, influence the attributed responsibility (Shaver 1985). In addition, different people may have different factors relevant to them, so the evaluation is also influenced by individual perceptions that do not change situationally, such as preferences, perceptions of justice/fairness, or the way decisions are made (actively or passively) (Konow and Schwettmann 2016, p. 95). As they (can) increase or decrease the assigned responsibility, they have to be taken into account.

It is easier to imagine alternatives to an action than to an act of omission, since no action is taken (Kahneman and Miller 1986, p. 136). Regarding the definition of causality, it is also questionable whether a decision-maker is causally responsible for an act of omission, since he has done nothing, i.e. no action has taken place. In addition, an act of omission may result from inertia (Thaler and Sunstein 2003, p. 177), inaction, or ignorance (Conrads and Irlenbusch 2013; Bartling et al. 2014a), which are not always easy to disentangle and which again call into question the causal relationship of omissions. For example, if the first decision-maker chooses the unequal allocation and the second decision-maker chooses the equal allocation, he may be perceived as deferring the decision to the third decision-maker (who is then inevitably the pivotal decision-maker) in order to avoid effective punishment (among others, see e.g. Bartling and Fischbacher 2012; Oexl and Grossman 2012). Alternatively, the choice of equal could be his preference, or the result of laziness to change the

default option. Either way, the intent behind an act of omission is less obvious (Spranca et al. 1991, p. 77), so judging or comparing fair and actual allocation is more difficult and may be inaccurate.

In this respect, assigning responsibility to the pivotal or initial decision-maker also depends on causality and counterfactual thinking. Each part of a decision sequence is evaluated in terms of its probability of changing the outcome, while responsibility is assigned in terms of its relative contribution. However, an act of omission does not change the probability of the outcome, so no (or less) causal responsibility is attributed (Spellman 1997, p. 345). This is true for the initiator, since it is imagined what might have happened if his action had not been taken or had been taken differently (Spellman 1997, pp. 323-327), but also for the pivotal decision-maker, where, i.e., in the Structural Model Approach, the minimum number of chances for an action to be counterfactually dependent on the outcome is considered (Chockler and Halpern 2004, pp. 94-95).

Additionally, when people assign responsibility, they consider the fairness of the situation, the outcome, and the process that leads to the outcome. Because different people apply different notions of fairness, there is no single understanding of fairness that applies to everyone (Cappelen et al. 2007, p. 818, 2010, p. 441). Consequently, fairness preferences are heterogeneous, pluralistic, and context-dependent, as different fairness principles are shared, but their interpretation and how they are valued depend on the context (Konow and Schwettmann 2016, p. 95). Again, a just or fair distribution may differ depending on the situational context, with people having different perceptions of fairness or justice in the same situation. Consequently, judgment and reaction to the same decision sequence may differ and may not be explained by a single criterion consistent with the multidimensional concept of responsibility attribution.

Typically, behavior that is inconsistent with self-interest and payoff maximization is often explained by social preferences. Accordingly, people are either inequality averse and prefer equal payoffs for all (Fehr and Schmidt 1999; Bolton and Ockenfels 2000), or focus on the intention of the decision-maker (Rabin 1993; Dufwenberg and Kirchsteiger 2004), or take outcome and intention into account when evaluating the fairness of an allocation (Falk and Fischbacher 2006; Falk et al. 2008). Each preference is associated with different expectations about how punishment is distributed. Inequality-averse individuals punish an unequal outcome because payoffs are not distributed equally, while equal outcomes are not punished. Intention-focused individuals punish when the decision reveals an unkind intention, i.e., when choosing the unequal allocation, though this may be mitigated by a passive implementation (an act of omission) because the underlying intention is not so obvious. Bringing the two

models together, an individual may consider the implemented outcome in addition to the (unkind) intention when deciding to punish.

Complementary to social preferences, distributional preferences specify (objective) criteria for how distributions should be allocated in a just and fair manner. Equality, consistent with outcome-based models of social preferences, postulates equal payoffs for all (Konow and Schwettmann 2016, pp. 84-85), while equity distributes resources according to the input provided (Becker 2012; Hülle et al. 2018). Consequently, a decision-maker who provides the same input as the recipient should have the same output. If not, punishment is likely. The need criterion distributes resources according to individual needs (Miller 1999), while the entitlement criterion specifies to allocate resources based on past performance or acquired characteristics with the entitled person having access to more resources mitigating the assigned punishment (Hülle et al. 2018). However, how these entitlements are (legitimately) allocated is not necessarily clear, as will be shown below. Accountability (or luck egalitarianism) distinguishes between controllable and uncontrollable factors (i.e., those that affect production), while an individual is not responsible for uncontrollable factors because he is not causally connected to them (Konow 2000, pp. 1073-1074; Dworkin 1981). Consequently, without a causal connection to an inequitable outcome, a decision-maker is not responsible and thus should not be punished.

The situational dimension of social norms is quite obvious for distributional preferences. In some situations, people may prefer an equal distribution and refuse to have more than others, while in other situations they may prefer to have more than others (Fershtman et al. 2012, pp. 133-140). Norms may also postulate the appropriate distribution of income or goods. In this regard, they specify what is perceived as a fair allocation, which is often a fifty-fifty split (Elster 1989, p. 101). Fehr and Fischbacher (2004) examine the egalitarian distribution norm in a dictator game and highlight that almost two-thirds punish if this norm is violated. The more the norm is violated, the more punishment is imposed (pp. 78-79).

Speaking of distributive justice and fairness, the second dimension comes into play: the procedure. Since the process can be the choice of an allocation (either active or passive) or the mechanism for assigning different status positions, as in this study, the justice or fairness of these processes is essential. Similarly, to judge whether a procedure is fair, or rather legitimate, counterfactual thinking takes place. Thus, the procedure actually used is compared to a fair procedure (Vermunt and Steensma 2016, p. 223). Consequently, the same relationship exists in the legitimacy assessment of the group-building process, which assigns decision-makers and recipients to their roles (visualized in the upper part of Figure 5.1).

In general, how positions or entitlements are assigned in laboratory experiments, either

randomly or through a specific procedure, does not automatically lead to significant differences in allocation decisions and reciprocal reactions (see e.g. Hoffman et al. 1994; Demiral and Mollerstrom 2018). However, usually people with high status positions are treated better and are seen as more deserving and claiming more for themselves (Ball and Eckel 1998; Albrecht et al. 2013). A status is a position based on a ranking in a society or group, where one position is simply seen as better than the other (Ball and Eckel 1998). In our experiment, we induce status differences by assigning participants to different status positions based on either a more legitimate random procedure or the final grade as less legitimate. When status positions are perceived as legitimate, they lead to social validity and stability, so inequalities are more likely to be accepted (Ridgeway 2001, pp. 257-258). Consequently, it may be socially acceptable to have more than others, and unpleasant decisions or outcomes are less likely to be opposed, if status differences, or the process by which these differences are implemented, are legitimate (Fershtman et al. 2012, pp. 133-140). What is seen as legitimate and leads to just status differences is based on the internalized norms and values of a society, while either the actor or the action (or both) can be legitimate (Kelman 2001, p. 55; Tyler 2001, p. 416).

The distinction between empirical and normative legitimacy helps to explain why an actor can be seen as (normatively) legitimate even though his decisions are not accepted because he is not legitimate in an empirical sense (Hinsch 2008). Since we have an objective and legitimate criterion to assign roles, the requirements for normative legitimacy are met in our setting. However, if the decision-maker is not perceived as legitimate in an empirical sense because he does not have subjective approval, neither the actor nor his actions will be accepted and supported. Consequently, his decisions, especially unequal decisions, are less likely to be accepted, or rather the difference between just and actual distribution is less likely to be accepted, so (more) punishment is expected.

Although many different procedures are typically used in laboratory experiments, it is still an open question as to what exactly constitutes a more (less) legitimate procedure for inducing status differences (we refer to Table 4.3 for an overview of different procedures). Furthermore, a procedure or mechanism may be legitimate for assigning roles in a dictator game, but not in every other decision situation. In order to find two mechanisms for our experiment, one of which is perceived as more legitimate and the other as less legitimate, we conduct a short pretest in which we evaluate four different mechanisms with respect to their legitimacy in our concrete decision context.

Before doing so, however, we want to summarize the determinants we identified for determining the just allocation, thus, the comparison between the actual and the just allocation

leading to the judgment of responsibility, which in turn determines the reaction, i.e., the decision to punish.

Summary 5.1: Attribution of Responsibility in Groups

The *decision-making process* is determined by the comparison between the actual and the just allocation, through the counterfactual consideration of what other allocation would have been possible or more just. Therefore, it, determines the judgment of responsibility that leads to punishment. The perception of what constitutes a just allocation is determined by situational social norms as well as individual social and distributional preferences. Defaults and acts of omission are perceived less negatively, and thus punished less, because the intention and causal role of the decision-maker is less clear.

In addition, in the *group-building process*, the actual procedure is compared to a (potentially) fair procedure, which determines the legitimacy judgment. Unkind decisions that result in inequalities are more likely to be accepted if status differences are legitimate, so less punishment is imposed.

Even with this theoretical clarity, the question of what constitutes appropriate behavior remains open. Consequently, it may lead to inconsistencies with prevailing social norms, as each individual may take a different norm as a starting point, leading to different evaluations and thus different reactions (Fershtman et al. 2012, pp. 133-140). In addition, the comparison of the actual allocation, outcome, or process with other (possible) alternatives also depends on the available allocations, outcomes, or processes, which are mentally constructed (Wells and Gavanski 1989, pp. 161-167; Malle et al. 2014, pp. 151-156). For example, unfair distributions in ultimatum games are more likely to be accepted if another unfair distribution could be chosen. Conversely, they are more likely to be rejected if another fair distribution is possible (Falk et al. 2003; Sobel 2005). Thus, counterfactual thinking also depends on availability.

There is another dimension that can influence the evaluation: the gender of the recipient. The presented experimental and empirical evidence suggests that women are more inequality averse (Andreoni and Miller 2002; Dufwenberg and Muren 2006), more loss averse (Schmidt and Traub 2002), and more risk averse (Holt and Laury 2002; Croson and Gneezy 2009) than men, which is expected to affect their punishment decision. Moreover, in a public goods

game, women are slightly more likely to punish (Burnham 2018), while men are more likely to support punishing unfair behavior (Singer et al. 2006). Consequently, women are more sensitive to the situational context of the decision situation, leading to greater behavioral variability, so the evaluation or comparison between actual and fair allocation may also differ between men and women. In addition, men are more sensitive to social hierarchies, which in turn may affect their punishment behavior, as they may (generally) punish more when status differences are induced (Burnham 2018).

By now it should be obvious that the attribution of responsibility is a multidimensional construct, with various relevant factors that depend on the individuals involved and the situational context (Shaver 1985). In the following second part of this study we come to our experiment with the purpose of shedding light on two dimensions and their implications on the individually assigned responsibility – more specifically, on how status differences and actively/passively implemented decisions determine the recipient's reaction.

Part II.

Experimental Design and Results

6. Experimental Design

The second part of our study presents the experiment, which aims to elucidate how responsibility is attributed to individual decision-makers in a group decision-making context with status differences and active (passive) decisions. A challenge, which we believe our design overcomes, is to investigate the (potential) interaction of a default and status differences on the attributed responsibility. Since the decision context has already been used in other experimental studies, comparisons with them are possible.

In our experiment, we first implement a default option with a preselected allocation that could be passively implemented. Since the effect of defaults implemented by doing nothing has already been studied in various experimental settings (among others, see e.g. Spranca et al. 1991; Jachimowicz et al. 2019), its expected direction of action is quite clear. Nevertheless, the purpose is to investigate whether a default also influences punishment behavior in our setting. Second, we induce more or less legitimate status differences between participants with the expectation that inequalities are more likely to be accepted if they result from a legitimate process (among others, see e.g. Albrecht et al. 2013; Hegtvedt et al. 2003).

To our knowledge, combining status differences with a default in such a setting has not been done before, so their combined effect on the assigned responsibility remains speculative. The experiments by Bartling et al. (2015) and Anselm et al. (2022) divide participants randomly to their roles, which may be perceived as less (more) legitimate, influencing the punishment behavior. Consequently, it is not clear whether recipients punish because they want to assign responsibility or because they evaluate the group-building mechanism as less (more) legitimate and (do not) accept the higher status of decision-makers.

However, before turning to our main experiment, we need to clarify which group-building mechanism is actually perceived as more or less legitimate. As we have seen in Section 4.3.2, there are many different procedures that are typically used to induce status differences, all of which have the same intention of placing participants into different roles or positions that are socially accepted. Since the perception of what participants consider to be a more (less) legitimate mechanism is the core element of the acceptance of status differences (Weiss and Fershtman 1998, p. 802; Bettencourt et al. 2001, p. 521), we want to make sure that we use two mechanisms that are actually perceived as legitimate (or less legitimate). Our pretest ensures exactly this aspect by testing four different mechanisms for their legitimacy against each other.

Therefore, Part II of this study is divided into three chapters that present the experimental design (Chapter 6), the results (Chapter 7), and conclusion of our study (Chapter 8). In this chapter, we start with the pretest (Section 6.1), where different group-building mechanisms are evaluated according to their legitimacy in inducing status differences. Then, the most and least legitimate mechanism is used in our experiment as a procedure for assigning different status positions.

In Section 6.2, we present our experimental design and the general procedures (Section 6.3) of the data collection process. Due to the proximity to the experiment of Bartling et al. (2015), Section 6.4 provides a brief methodological discussion of the two different elicitation methods, the direct response and the strategy method, as well as an overview of the quantity and quality of replications. This chapter concludes with the hypotheses (Section 6.5), which are the contextual transition to Chapter 7, where the results of the experiment are presented.

6.1. Pretest

6.1.1. Evaluation of Different Group-Building Mechanisms

As shown in the previous chapter, there are various ways to assign participants in laboratory experiments to groups that occupy different roles or positions. Acceptance of the status differences created in this way is relevant and influences individual behavior and the acceptance of decisions and their resulting outcomes. Thus, status structures must be perceived as legitimate in order to be accepted, as this may determine the attribution of responsibility (among others, see e.g. Turner and Brown 1978; Bettencourt et al. 2001; Olson and Hafer 2001).

However, there is no clear concept in the experimental literature of what exactly constitutes a legitimate or illegitimate group-building mechanism. Because this is fundamental to the primary research idea of this study, a brief pretest is administered in which four group-building mechanisms are tested on a scale of 1 to 7 in terms of how legitimate they are. Then, the most and least legitimate mechanisms are used in the laboratory experiment to actually divide participants into decision-makers and recipients. The four mechanisms, random, slider, Raven's matrices (IQ), and final grade, represent the four main categories of procedures presented in Table 4.3.

Before turning to the design and results of the pretest, each mechanism is explained in detail[42] and assumptions about the legitimacy rating are formulated.

[42]See Appendix A.1 for the instructions used in the pretest.

- **Raven's matrices:** Raven's matrices[43] are non-verbal multiple-choice questions designed to measure cognitive ability. Participants have a fixed time limit to select one of eight possible symbols/pictures that fit the pattern. The questions are adapted from Séréville and Myers (1994) and used by Putterman et al. (2011) and Paetzel and Sausgruber (2018). Since mathematical or knowledge quizzes only test ability in a specific area, the matrices are more general and do not require advanced school or college education.

- **Final grade:** The final (school) grade is a characteristic acquired in the past and has no direct relation to the situation in the laboratory. The effort or ability required to achieve the grade is not measurable during the experiment. People who have a 'bad day' or are unfamiliar with the experimental sessions are not disadvantaged by this mechanism, while participants with laboratory experience have no advantage. However, the associations with school grades are quite polarizing. On the one hand, school grades should provide information about the individual's intelligence, ability, and knowledge (Wentzel 1991; Camara 2005). On the other hand, school qualifications often depend on many aspects, such as the family situation (parents' education) or the school itself (private vs. non-private), or the local place (city or urban area) (Süß 2001; Ingenkamp 1997). Thus, it is controversially discussed what the final grade might actually reflect.

- **Slider task:** The slider task represents real effort tasks that measure effort within a short and predefined time. On a (web) page, sliders are displayed at a random number between 0 and 100, and participants have to move them to a given (also randomly determined) number. The goal is to complete as many sliders as possible within a given time limit. Therefore, this task is easy to understand and does not require any special knowledge or skills, although a keen perception or experience in using a computer mouse may be helpful (Gill and Prowse 2019, pp. 2-3).

- **Random:** Random assignment is the most common and standard way to assign different roles or positions. Typically, when status differences are implemented, randomly assigned roles are used as the baseline, since no special skills, knowledge, effort, or other (external) characteristics determine them (among others, see e.g. Hoffman et al. 1994; Bartling et al. 2015). Thus, all participants have the same probability of being assigned to any position or status group.

Based on empirical legitimacy, we expect that Raven's matrices will be evaluated as the

[43]Hereinafter the term IQ is used to refer to this mechanism.

most legitimate mechanism because it signals (cognitive) ability and thus provides concrete information about the status holder. Although the ability is not related to the task, it may be perceived as a good measure of competence. Regarding the evaluation of the slider task and the final grade, we expect that the final grade will be perceived as more legitimate than the slider task. As with the Raven's matrices, a good final grade can signal competence and ability, although it is less neutral than the Raven's matrices because grades in school are always somewhat subjective. The slider task does not require any special knowledge or ability, only effort, so it provides less information about an individual with many correct sliders. Finally, since a random procedure provides no information and is completely arbitrary, we expect it to be rated as the least legitimate mechanism.

6.1.2. Design and Procedures

The pretest opened with an explanation of the hypothetical situation and the different group-building mechanisms that the participants had to evaluate. First, the situation itself was explained to the participants. They were told that a group of six people would be divided into three decision-makers and three recipients. The three decision-makers then had to divide an amount of money between themselves and the three recipients. Participants were asked to rate the legitimacy of the four mechanisms for dividing the six people into decision-makers and recipients. For a better understanding, they were shown a picture that outlined the situation. After the context was explained, the four different mechanisms were presented in a randomized order. First, each mechanism was described in detail and, if relevant, a short example or picture was shown. Then, at the bottom of the page, a scale from 1 to 7 was displayed, with 1 being completely not legitimate and 7 being completely legitimate. Figure 6.1 shows a sample screen from the slider task. The scale at the bottom of the page was similar for all four mechanisms. After each mechanism was evaluated, a short questionnaire asking for demographic information was implemented.[44]

The pretest was programmed in oTree (Chen et al. 2016) and conducted online in July 2020 with students recruited via hroot (Bock et al. 2014). We used a within-subject design where all participants had to rate each mechanism. A total of 105 students participated, of which 36% are male; the mean age is 25.7 years.

[44] For the exact wording and instructions of the hypothetical situation and the four mechanisms, see Appendix A.1.

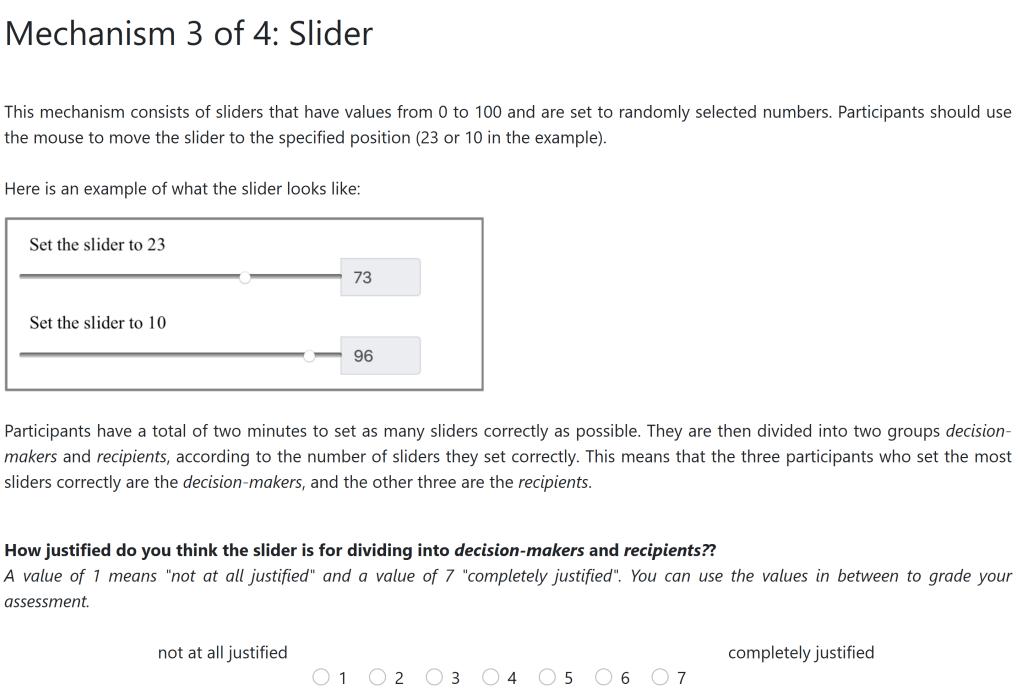


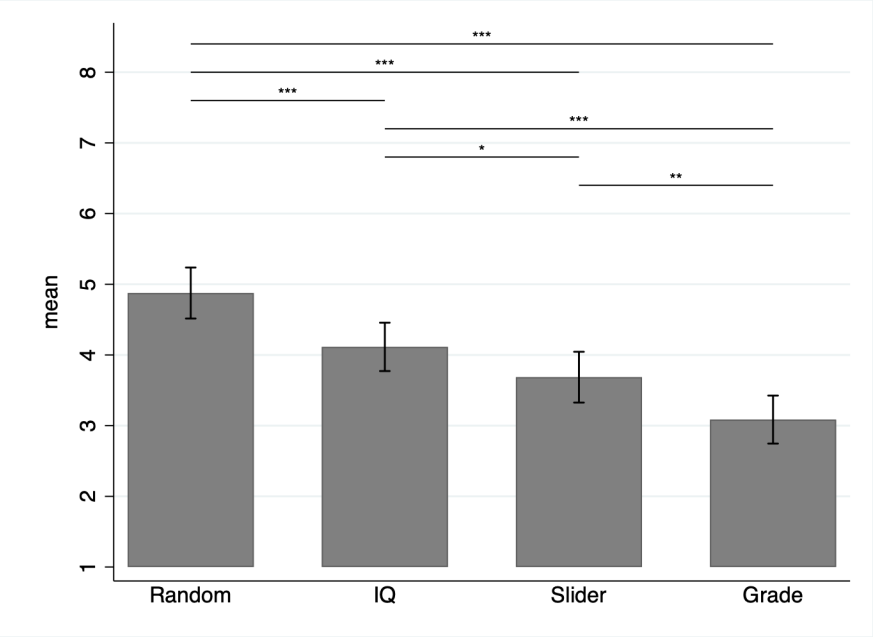
Figure 6.1.: Example: Slider

6.1.3. Results

This section presents the results of the pretest and their implications for the upcoming experiment. First, we present descriptive statistics and look at gender differences. We then turn to tobit regression analyses. For this purpose, the data is transformed into a panel structure, where each mechanism represents a time component, resulting in four observations per participant.

Recall that the rating for each mechanism ranges from 1 (= completely not legitimate) to 7 (= completely legitimate). Figure 6.2 shows the mean rating for each mechanism, with larger numbers indicating higher legitimacy. Error bars represent the 95% confidence intervals, and p-values are calculated using two-tailed, within-subjects t-tests.[45] With a mean of 4.88, a random procedure is rated as the most legitimate group-building mechanism, while the final grade is rated as the least legitimate mechanism (mean of 3.09). All mechanisms differ significantly from each other, at least at the 10%-level (IQ and slider). The evaluation of random as the most legitimate mechanism is further supported by the fact that the difference to the three other mechanisms is highly significant at the 1%-level.

[45] All t-tests presented in this study refer to the two-tailed significance level unless otherwise noted. In addition, nonparametric rank sum tests are computed and mentioned when they result in different levels of significance.



Notes: Rating scale range from 1 (completely not legitimate) to 7 (completely legitimate). The p-values are from two-tailed t-tests, assuming equal variances. Bars represent the 95% confidence interval. $^{*}p \leq 0.1$, $^{**}p \leq 0.05$,$^{***}p \leq 0.01$.

Figure 6.2.: Mean Rating of the Four Mechanisms

As mentioned in Section 3.2, men are generally more competitive than women and more sensitive to (status) hierarchies (among others, see e.g. Campbell 2002; Burnham 2018), so it is valuable to discuss the ratings separately for men and women. Table 6.1 provides the results of this comparison.

The overall ranking of the four mechanisms and which is considered most/least legitimate does not differ between men and women. However, a closer look at the results indicates that women are more extreme in their evaluation behavior than men. Specifically, the spread between the most legitimate procedure (random) and the least legitimate mechanism (final grade) is 2.2 points for women and 1 point for men. Women rate a random procedure as almost twice as legitimate as the final grade (5.01 vs. 2.81). Thus, a random procedure is more legitimate for women than for men, while the final grade is less legitimate for women than for men (the difference is significant at the 5%-level).[46] The rating scheme for male participants shows that the differences between the mechanisms are less pronounced. Specifically, there are only (weakly) significant differences between random and final grade (5%-level) as well as random and slider (10%-level); there are no other significant differences. Except for IQ

[46] A corresponding rank sum test shows a significance level of 10%.

Table 6.1.: Mean Rating of the Four Mechanisms

Rating	Men	Women	t-test
Random	4.61	5.01	$p = 0.295$
	(0.322)	(0.220)	
IQ	4.33	4.00	$p = 0.361$
	(0.295)	(0.213)	
Slider	3.86	3.59	$p = 0.487$
	(0.301)	(0.227)	
Grade	3.61	2.81	$p = 0.026$
	(0.336)	(0.186)	

Table notes: Rating scale range from 1 (completely not legitimate) to 7 (completely legitimate). Standard errors in parenthesis. The p-values are from two-tailed t-tests of differences between men and women, assuming equal variances.

and slider, women rate all mechanisms differently at the 1% significance level.

Next, we turn to more advanced regression analyses that examine the (possible) influence of gender and other sociodemographic characteristics. Table 6.2 reports the results of a tobit regression where the rating (from 1 to 7) is the dependent variable and a random mechanism is the base category.

Since R^2 is not valid for panel data tobit regressions, we use the Akaike's Information Criterion (AIC) values to assess which estimated model is the best. The AIC value measures the combination of model fit and complexity, where fit is measured negatively, meaning that the smaller the value is, the better the fit is (Akaike 1974).[47] To compare the AIC value(s) of our models, we compute them for each estimated model and then rank them, with the model with the smallest AIC-value being the best model (Claeskens and Jansen 2015, p. 648).

Table 6.2 confirms the results already highlighted. In Model (1), all three mechanisms are significantly different from a random procedure. Controlling for socio-demographic charac-

[47]There is a second measure, Bayesian Information Criterion (BIC), that is calculated similarly and is also commonly used to compare different models. Both values measure the fit of the model equally, while the complexity differs between the two values. The AIC doubles the number of parameters, while the BIC multiplies the number of parameters by ln of the number of observations. Specifically, AIC and BIC are defined as follows: $AIC = -2 \cdot ln(likelihood) + 2 \cdot k$; $BIC = -2 \cdot ln(likelihood) + ln(N) \cdot k$, where k = number of estimated parameters and N = number of observations. Consequently, the two values differ in their calculation of complexity. Determining the useful value of N is problematic for BIC, especially for dependent observations (Akaike 1974; Raftery 1995). Since this may be relevant to our data, we use the AIC values to compare our models.

Table 6.2.: Tobit Regression - Pretest

Rating	(1)	(2)	Only male	Only female
IQ	-0.762***	-0.780***	-0.125	-1.019***
	(0.235)	(0.236)	(0.422)	(0.282)
Slider	-1.190***	-1.194***	-0.756*	-1.422***
	(0.235)	(0.234)	(0.408)	(0.282)
Grade	-1.790***	-1.793***	-0.960**	-2.211***
	(0.235)	(0.235)	(0.416)	(0.281)
Female		-0.208		
		(0.211)		
Age		-0.015	0.034	-0.027
		(0.020)	(0.048)	(0.021)
Riskneutral		-0.368	-0.657	-0.150
		(0.282)	(0.442)	(0.359)
Riskaverse		-0.458*	-0.430	-0.472
		(0.247)	(0.416)	(0.303)
Second round		-0.091	-0.503	-0.025
		(0.235)	(0.420)	(0.282)
Third round		-0.261	-1.024**	0.086
		(0.234)	(0.410)	(0.282)
Fourth round		0.063	-0.354	0.199
		(0.236)	(0.419)	(0.281)
Constant	4.876***	5.811***	4.529***	5.942***
	(0.176)	(0.613)	(1.333)	(0.680)
Wald-χ^2	61.52	69.34	17.11	68.51
$p(\chi^2)$	0.000	0.000	0.047	0.000
AIC	1692.54	1699.28	599.57	1107.15
N	420	420	144	276

Table notes: Random effects tobit regression with panel estimator. Rating scale from 1 (= completely not legitimate) to 7 (= completely legitimate) as dependent variable. Female is a dummy variable for the respective socio-demographic variable. Random group-building mechanism, riskaffine and first round as baseline categories. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

teristics (Model 2) strengthens this result. On average, the rating of IQ is 0.78 points lower than the average rating of random. The slider task is rated 1.19 points less legitimate and the final grade is 1.79 points less legitimate than random ($p < 0.01$). In addition, the self-rated risk attitude has a significant effect on the rating, as risk-seeking people rate on average 0.46 points higher than risk-averse people ($p < 0.1$). Comparing the AIC of the two models shows that the former has a lower AIC and is therefore better than the latter. However, since we are only using these analyses to justify the most and least legitimate mechanism, it is not of primary interest which model fits best.

Since Table 6.1 revealed gender-specific rating differences, we split our dataset and estimate the tobit regression separately for men (Model 3) and women (Model 4). Looking at Model 3, we see that men do not rate IQ as significantly more legitimate than a random procedure. However, the slider task and final grade are found to be more legitimate, although the effects are only statistically significant at the 10%- and 5%-level, respectively. Men seem to rate the mechanism presented in the third round as different from the mechanisms presented in the first round. Since the distribution of mechanisms presented in the two rounds is quite similar, this effect seems to be an artifact of the experimental design.

The final model includes only women. As expected from the previous analyses, the evaluation behavior of women is more dispersed. More specifically, the difference between the most legitimate mechanism random and the least legitimate mechanism final grade is 2.21 (out of seven), which is highly statistically significant ($p < 0.01$). In addition, IQ and slider are also considered significantly less legitimate than random. Interestingly, the significant effect of risk attitude in Model 2 disappears when only men (or women) are considered. We can only speculate as to the reasons why risk-averse individuals generally evaluate less legitimately than risk-seeking individuals. To our knowledge, there are no experimental or empirical studies that examine the effect of risk attitude on valuation behavior, i.e., whether risk-averse people generally value differently than risk-seeking people.[48]

Table A.1 in Appendix A.2 reports the marginal effects of the last three tobit regressions and the mutual tests for significant differences (Wald-F-tests) between each mechanism. These analyses support the findings presented here that men's ratings are less dispersed and extreme in that not all mechanisms are significantly different from each other. Before turning to the implications of the pretest, the following summary highlights the two main findings.

[48]We will come back to this point later when we analyze the results of the main experiment.

Summary 6.1: Results Pretest

From the pretest we can conclude two important things. First, a *random* procedure is the most legitimate group-building mechanism, while *final grade* is the least legitimate. Second, the evaluation of the most and least legitimate mechanisms does not differ between men and women, although women evaluate the four mechanisms more extremely than men do.

6.1.4. Implications

The main implication of the pretest is that the two different group-building mechanisms we use in the main experiment are a *random* procedure and the *final grade*. With these two we divide participants into high and low status groups. However, before turning to the main experiment, this section provides a brief discussion of both mechanisms, starting with final grade and educational background in Germany, followed by randomization preferences in individual choices.

Education as an indicator for assigning people to different status groups was discussed in Section 4.1.1. Since education, or rather the final grade, is a crucial determinant for inducing status differences in this setting, we discuss this topic in more detail here. When personal characteristics such as gender or education are associated with a person's status, these characteristics are often used to infer the person's performance (Ridgeway et al. 1998; Fleiß 2015). How status is acquired, either through individual characteristics or through actions, is quite important. Weiss and Fershtman (1998) point out that in the US school education is the key level for attaining higher status (pp. 805-806), which is likely to be true for almost all countries. Accordingly, a good education is in the interest of many people.

In the 2020/2021 German school year, more males than females graduated from high school, while more females than males obtained the Abitur (55.15% of all graduates with the Abitur were females). Of all female graduates, 38.73% obtained the Abitur, while 30.06% of all male graduates left school with the Abitur (Bundesamt 2022). The final grades of the Abitur differ depending on the geographical region, since the school system is the responsibility of the state and not the federal republic. The average final grade was lowest in Thüringen (2.06) and highest in Rheinland-Pfalz with an average of 2.48 (Konferenz 2021).

In addition to obtaining a high level of education, the final grade plays an important role. Usually, grades depend on individual performance and are important when applying for college or a job. When talking about education or qualifications, performance is often used

as a predictor of how professional positions are distributed or promotions are awarded (Fleiß 2015, p. 390; Burnham 2018, p. 2). As Wenzel points out: "Grades reflect learning that takes place within the larger social context of the classroom and that requires effort and persistence over long periods of time. In contrast, other measures of achievement, such as performance on standardized tests, assess basic or specialized academic abilities and aptitudes at one point in time without social influence" (Wentzel 1991, p. 1068).

In a cross-national meta-analysis of gender differences in school achievement, Voyer and Voyer (2014) reveal that women have better grades in most courses at school, although they do not perform better on achievement tests where men generally do better, especially in math and science. Women's better school performance exists at all levels of schooling: elementary school, middle school, high school, and college (p. 1174). Intelligence and academic success are positively correlated (r between 50 and 60), although other factors such as motivation, cognitive ability, school-context and home context also determine academic success. One explanation for why intelligent people do better in school is that they can process information more easily and quickly (Fritz et al. 2018, pp. 171-172).

In Germany, the criteria for selecting students for universities is determined by the 'Hochschulrahmengesetz'. According to this, the final grade is mandatory for the selection process, while other criteria such as weighted individual grades, tests, interviews, or previous work time (such as vocational training) are optional. In this context, the final grade is the best indicator of success in studies (among others, see e.g. Brandstätter and Farthofer 2002; Gold and Souvignier 2005). Heine et al. (2010) use a questionnaire design and a representative sample of graduates of the year 2008 in Germany to investigate what people with a higher education entrance qualification (generally the 'Abitur') do half a year after graduation. The majority of them, 62%, have already started their studies or vocational training (p. 1). Additionally, they use logistic regression to analyze potential factors in the decision to study, using the mean grade of all participants (2.3) as the benchmark.[49] They find a significant effect of final grade on the likelihood of going to college. After controlling for other factors such as gender and social background, a drop of 1 grade reduces the probability of enrolling by 17 percentage points (Heine et al. 2010, pp. 52-54).

However, there is also criticism of the final grade. On the one hand, the final grade depends on a two-year period at school, where the grade is based on many exams in different courses, so that aspects such as work effort at school and conscientiousness are considered (Camara

[49]Not surprisingly, the mean grade differs by family background. Graduates whose parents have a university degree have an average grade of 2.18, while graduates whose parents do not have a university degree have an average grade of 2.38, which is significantly different at the 1%-level (Heine et al. 2010, p. 50).

2005). On the other hand, grades are also determined by subjective comparison criteria and other non-academic determinants. As a result, there are (sometimes) large differences between classes, courses, and even states, because grades are subjective evaluations of teachers (Süß 2001; Ingenkamp 1997).

Much of the research on status-related hierarchies has focused on SES, i.e., social class, which refers to the economic or educational standing of one person relative to another (among others, see e.g. Piff et al. 2010). However, relative educational attainment within the same educational group, as expressed by final grade, has not been studied in experimental settings. To our knowledge, there is no experimental study in which participants are ranked according to their final grade and then divided into groups. As stated by SCT, the final grade (as a form of education) can be seen as a diffuse status characteristic. Therefore, it does not implement specific performance expectations in the task, as no special skills or abilities associated with school performance are required. However, even the diffuse character of the final grade can form expectations for the following group decision (Berger et al. 1966).

In our experiment, the final grade has nothing to do with the subsequent task and is not an indicator or predictor of a (more) successful decision. Therefore, participants may perceive the final grade as empirically unrelated to the task and somehow random. As presented here, grades are seen as subjective and influenced by other criteria. In Germany, however, they are always (and sometimes exclusively) used to select students for study programs because they are a good indicator of success.

Somewhat unexpectedly, participants rated randomization as by far the most legitimate mechanism for assigning people to different positions. However, there is a strand of literature that can explain such a rating behavior. In this context, there is a body of (experimental) literature showing that people do not (always) have stable preferences, as they (randomly) make different choices even when options are the same, which contradicts the completeness axiom as a fundamental assumption in economic decision making (Cettolin and Riedl 2019). There are three prominent explanations for such stochastic or random choices. (1) People make stochastic choices because their preferences are stochastic. (2) People make stochastic errors because they have stable preferences. (3) People voluntarily change their choices to achieve an optimal level of outcomes or to minimize regret (Agranov and Ortoleva 2017, p. 41, 2022). Specifically, people are found to have a preference for randomization in that they prefer a random procedure (such as a coin flip) to choose for them, especially when their decision affects another party. Cettolin and Gächter (2019), for example, find that half of subjects behave inconsistently with expected utility models that assume complete

preferences. In addition, when a decision determines one's own payoff and the payoff of another party, Sandroni et al. (2013) show that nearly a third of participants have a tendency to implement such a decision through a random device rather than choosing it themselves. Even if such a randomization procedure is costly, people show a tendency to delegate their decision (Agranov and Ortoleva 2017).

Consequently, people have been found to have a preference for randomization in some situations (among others, see e.g. Sandroni et al. 2013; Cettolin and Riedl 2019; Agranov and Ortoleva 2022), which may explain why they rate a random group-building mechanism as the most legitimate procedure. In addition, because a random procedure is completely arbitrary, each participant has an equal chance of being assigned to each role, which may have its own value for participants.

6.2. Experimental Design

Now that we have established the two mechanisms that induce status differences, we can turn to the experimental design. Therefore, this section gives an overview of the treatments and explains each part of the experiment in detail.[50]

Combining the two default options (equal and unequal allocation) with the two group-building mechanisms (random and final grade) results in a 2x2 design (see Table 6.3).

Table 6.3.: Treatment Classification

		Group-Building Mechanism	
		Random	Grade
Default	Unequal	random-unequal (T1)	grade-unequal (T2)
	Equal	random-equal (T3)	grade-equal (T4)

The experiment consists of four parts, which are described in detail in this section. The first part is a simple dictator game with punishment, followed by the group-building stage in the second part. The main experiment consists of the third and fourth parts, the sequential dictator game with punishment, where the basic structure is adapted from Bartling et al. (2015). However, in contrast to Bartling et al. (2015), we do not use the strategy method but

[50]Pretest participants are excluded from the laboratory experiment.

the direct response method and repeat parts three and four for six rounds. At the end of the experiment, participants are asked to complete a short questionnaire. Each part is described in detail below.[51] We have used neutral language in the instructions, avoiding terms such as *punishment* or *status*.

Figure 6.3 gives an overview of the experimental design highlighting the essential facts of each of the four parts. The two treatment variations (group-building mechanism and default) are highlighted in red.

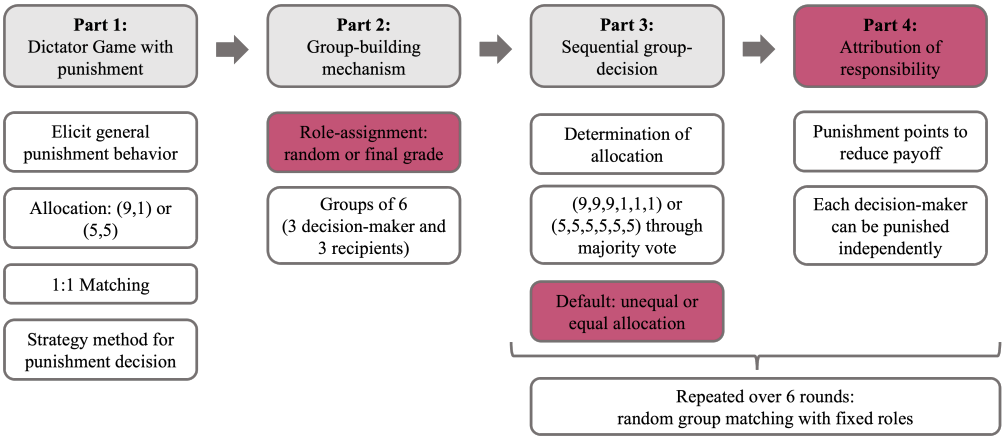


Figure 6.3.: Overview of the Experimental Design

Part 1 - Dictator Game with Punishment

The experiment starts with a simple dictator game with punishment using the strategy method. Therefore, participants play in pairs, each deciding as a dictator and then as a recipient. The dictator chooses between two allocations that divide points between himself and the matched recipient. One allocation is equal and divides 5 points between the dictator and the recipient; the other allocates 9 points to the dictator and 1 point to the recipient. Then, all participants act as recipients and announce whether they want to punish and how many punishment points (maximum of 5 points) they assign to each of the two choices. After deciding as dictator and recipient, participants are randomly assigned to one of the two roles. The decision they both make in their assigned role is implemented and determines the payoff for the first part, although this information is not made public until the end of the experiment.

[51] During the experiment, parts 1 and 2, and 3 and 4 are treated together in that the instructions for parts 1 and 2 (and 3 and 4) are read aloud together.

Part 2 - Group-building Mechanism

Before starting the sequential dictator game with punishment, each participant reports the final grade of his 'Abitur' (= high school diploma) or the equivalent grade required to enter university. Even in the randomized treatments, the final grade is queried in order to make the treatments as similar as possible. Then, between parts two and three, participants are assigned roles, with role A being the decision-maker and role B being the recipient. In the following experiment, the letters A and B are used to avoid framing the participants. In the two random treatments, the roles are randomly assigned, whereas in the grade treatment, the median grade of a session is calculated. Participants with a score above the median are designated as decision-makers, while all below the median are designated as recipients.[52] The instructions explicitly explain how participants are divided into their roles, and that participants above the median become decision-makers, while the others are recipients. No specific information about the exact rank in the session is provided, and the assigned roles are constant for the rest of the experiment. This procedure of role assignment is not commonly known until the second part of the experiment.

Part 3 and 4 - Sequential Dictator Game with Punishment

At the beginning of the third part, the participants are informed about their role and are instructed about the following decision context. After the instructions are read, participants must answer three comprehension questions. The following decision situation consists of two stages: (1) the sequential group decision, which determines the allocation, and (2) the individual allocation of punishment points as a measure of responsibility.

In the third and fourth parts of the experiment, participants are divided into groups of six, each with three decision-makers and three recipients. After each round, the groups are randomly reassigned, with players remaining in their roles (either as decision-makers or recipients). However, they form a new group with participants different from those in the previous round(s). At the beginning of each round, players are informed of their position in the group, which determines when they have to decide. The first decision-maker has position A1, the next decision-maker has position A2, and the last decision-maker has position A3. The next steps, 1 and 2, are repeated for six rounds.

Stage 1: First, the three decision-makers decide, one after the other, which of two possible allocations they want to implement for the whole group. They have 30 points to distribute

[52] If there are multiple participants with a score equal to the median, they are equally divided between the two roles, so each role contains the same number of players.

among their group members, for which there are two allocations. Allocation 1 (5, 5, 5; 5, 5, 5) gives each group member the same number of points, while allocation 2 (9, 9, 9; 1, 1, 1) distributes 9 points to each of the three decision-makers and 1 point to each recipient. Depending on the treatment, an allocation is preselected that remains the same for the entire session. In addition, each decision-maker has 30 seconds to click the 'next button' to confirm their choice. If the button is not clicked within 30 seconds, the marked allocation is implemented (this could be the preselected allocation or the other if the decision-maker switches to the other but does not confirm his choice). The outcome is certain if at least two decision-makers choose the same allocation. The decision-making process is transparent, so that the decisions of each decision-maker can be observed. When the first decision-maker makes his decision, the second decision-maker is informed before he makes his decision. Similarly, the third decision-maker observes the chosen allocation of decision-makers 1 and 2. The decision situation is also transparent for the recipients, as they see the decision of each decision-maker in their group. During the sequential decision phase, the recipients are asked about their expectations. At the beginning of each round, they are assigned to a position (B1, B2, or B3) and participate in the same sequential decision context. They indicate what decision they expect from the corresponding decision-maker in the same position. Specifically, receiver B1 states his expectation of A1's decision, B2 states his expectation of A2's decision, and B3 states his expectation of A3's decision.

Stage 2: After the outcome is determined, the recipients can assign punishment points to each decision-maker in their group individually. To do this, they see the implemented allocation, the default allocation, and the allocation chosen by each decision-maker. They can then assign up to seven punishment points at the cost of one point, using the assigned punishment points as a measure of responsibility. Regardless of the number of punishment points assigned, the payoff for that round is reduced by one point if at least one punishment point is assigned. The seven punishment points can be assigned to a single decision-maker or to two (three) different decision-makers. The only restriction is that no decision-maker can receive more punishment points than his payoff in this round, so a reduction below zero points is impossible. Meanwhile, the decision-makers are asked to state their punishment expectations. For this purpose, the implemented allocation, the default option, and each allocation chosen by the decision-makers in their group are displayed, and they indicate how many punishment points, from zero to seven, each decision-maker is likely to receive for his decision.

To determine the payoff of this part, a round and a receiver are randomly selected. The

payoff is determined by the allocation made minus the punishment points assigned in that round. Players do not know which round or recipient is relevant for payoff. After the sixth round, all participants are informed about the payoff of the experiment and are directed to the questionnaire.

Questionnaire

Apart from some socio-demographic characteristics (age, gender, study, semester), participants are asked to self-assess their risk attitude on a 5-point Likert scale. Afterwards, an open question is displayed in which the participants, either as decision-maker or as receiver, indicate their decisive reason(s). Additionally, depending on the group-building mechanism (random or grade), participants indicate the legitimacy of the two mechanisms on a 7-point Likert scale. This question is adapted from the pretest in order to make the results comparable. Furthermore, according to Demiral and Mollerstrom (2018), participants state the deservingness of the two roles A (= decision-maker) and B (= recipient). Finally, the short version of the BSJO-scale, as used in Hülle et al. (2018) and Adriaans et al. (2019), is displayed.

Using the dictator game in this setting is an excellent way to measure individual and group behavior because it is a simple and well-understood experimental game. In its simplest form, a dictator must decide how much of an endowment to give to another anonymous subject. Thus, the dictator game is about the distribution of resources or money and has been widely used in experiments (Engel 2011, p. 593).

Decision-makers have a limited 30-second window to actively choose, with one option, either the equal or unequal allocation, preselected. The selected option is implemented if participants do not click the 'next-button' within 30 seconds. This does not necessarily implement the default, as participants can switch to another option but not actively continue. The role of time pressure on decisions has been investigated in some recent studies (see e.g. Cappelen et al. 2015; Gärtner 2018). Although the maximum decision time is limited in our study, no direct time pressure is induced because the decision time is long enough to make an active decision and to think about it appropriately. Participants have enough time to read the instructions before they start. In addition, the time limit is used only for the decision phase, not for the responsibility attribution part.

Gärtner and Sandberg (2017) also implement a fixed time (40 seconds) for participants to actively decide, with an average decision time of 22 seconds. Furthermore, they use a questionnaire to assess whether the time limit is too short, highlighting that more than 95%

of participants agree with the time limit and find it sufficient (p. 13). Thus, the 30 seconds we used seems to be sufficient to allow participants to make an appropriate choice. Moreover, our study does not focus on the decision process, but on the evaluation of recipients and the availability of a default and passive implementation; therefore, mere effects of time pressure are negligible.

6.3. General Procedures

All sessions were conducted at the University of Hamburg between September and November 2020. For each of the four treatments, three sessions were conducted with either 18, 24, or 30 participants (as the group size is six, there had to be a number divisible by six). Each participant could only take part once, and all participants who took part in the pretest were excluded from the laboratory experiment.

A total of 270 participants, mostly undergraduate students from various disciplines, were recruited using the software hroot (Bock et al. 2014), resulting in 2430 punishment observations (three recipients per group over six rounds). The experiment was programmed using oTree (Chen et al. 2016). On average, a session lasted 60 minutes and was incentivized with 14.83 Euros, including a show-up fee of 5 Euros. As in Bartling et al. (2015), points were used instead of real currency, with a conversion rate of 1 to 1. Thus, 1 point was equivalent to 1 Euro, and participants were not informed about their payoff until the end of the experiment.

At the beginning of each session, the printed instructions for parts 1 and 2 were distributed to the participants and read aloud by the experimenter to ensure that the experimental design and procedures were familiar to all participants. After reading the instructions, participants were given the opportunity to ask questions. Prior to parts 3 and 4, the instructions for that part were distributed and read aloud, and time was allowed for questions to be asked.

6.4. Experimental Method

Conducting a laboratory experiment always raises questions about the method used to elicit participants' behavior. There are two different standard methods: the *direct response* method and the *strategy* method. From a game theoretical perspective, they should elicit the same behavior. However, the paper by Bartling et al. (2015) uses the strategy method to induce punishment behavior, while our study uses the direct response method. Therefore, both methods will be discussed in the following section by presenting methodological differences, advantages and disadvantages, and implications. Additionally, an overview of (experimental)

evidence and whether they induce behavioral differences is provided. Since our experimental design is adapted from Bartling et al. (2015) and replicates their experiment, considerations about replication of previous studies in experimental economics are appropriate. Therefore, a brief overview of replications is provided in Section 6.4.2.

6.4.1. Elicitation Procedures

The direct response method means participants decide when it is their turn. Consequently, the information gathered is limited to the specific decision situations and no alternative scenarios are considered. In contrast, the strategy method, developed by Selten (1967), involves decision-making conditional on the decision(s) of other participants for each possible set of alternatives. Thus, a complete strategy of the game is elicited, and data are collected even for rare situations (Brandts and Charness 2011, pp. 375-376). However, participants are asked hypothetically how they would decide given the situation presented, which encourages counterfactual thinking about other possible outcomes (Rauhut and Winter 2010, p. 1184). The strategy vector method is a further extension of the strategy method, where a complete strategy profile for the whole game is elicited without knowing one's own role. Thus, in asymmetric games, participants must decide for each possible situation, first in role A and then in role B, conditional on all possible decisions someone in role A could have made (Oxoby and McLeish 2004, p. 400; Rauhut and Winter 2010, p. 1184).

One could argue that using the strategy method is an abstraction of natural situations, encouraging participants to think differently about each decision. Using the direct response method involves (usually) a single decision at a time. Since the strategy method asks for a whole set of alternatives, participants may feel compelled to behave consistently and thus differently than they would in a single decision. As a result, the strategy method is seen as more rational and is often described as *'cold'*, whereas the direct response method is more *'hot'* and thus emotional (Roth 1995, p. 323; Brandts and Charness 2000, p. 228; Oxoby and McLeish 2004, p. 400).

A major difference between the two methods is the time required to complete an experiment. Typically, the strategy method involves more decisions and takes more time to complete. This must be taken into account, especially in laboratory settings where time is an important factor in determining participants' payoffs. However, measuring a complete game strategy has the great advantage of conditionality. Whenever a conditional dimension (i.e., social norms) is of interest, it allows to elicit behavior that may differ with the situational context (Rauhut and Winter 2010, pp. 1183-1187). In the first part of our exper-

iment, the strategy (vector) method elicits the general punishment behavior. It is necessary to derive this information conditional on all possible situations in order to discover the relatively complex dimension of the underlying norms. However, in the third and fourth parts of our experiment, the direct response method is appropriate because the time constraint is impossible with the strategy method.

Beyond the methodological differences and advantages of the two methods, it remains unclear whether they induce significant differences in behavior. Although research is limited, some empirical evidence attempts to explore this question further. The most comprehensive work to date on methodological differences is a review of 29 experimental studies by Brandts and Charness (2011). They analyze the experimental results of both methods and examine whether they induce different behavioral patterns. Overall, 4 out of 29 experiments show behavioral differences between the two methods, 16 find no differences, while nine show mixed results (for an overview of all experiments, see Brandts and Charness 2011, pp. 388-390). Since punishment decisions are of primary interest in our study, the results relevant to the analysis of punishment differences are discussed below.

Using a signaling game, Brandts and Charness (2003) prove that the direct response method doubles punishment rates compared to the strategy method. Another experiment emphasizes that first movers show more selfish behavior with the direct response method as long as the punishment is not costly (Brosig et al. 2003). In a prisoner's dilemma with the possibility of punishment, Falk et al. (2005) find that the probability of punishment does not differ between the two methods. However, when a participant does punish, the amount is almost twice as high in the direct response method as in the strategy method. Using an ultimatum game to examine offer and rejection rates, Rauhut and Winter support these results (2010). Participants play the game with both methods without receiving feedback between the parts. As in Falk et al. (2005), the offers per se do not differ; however, the acceptance rate is significantly higher in the direct response game than in the strategy game. The experiment by Oxoby and McLeish (2004) has partly mixed results. Using an ultimatum game, the mean offers of the proposers and the acceptance rates of the recipients do not differ between the elicitation methods. However, for small offers (3 or less out of 10), rejection rates are higher with the direct response method than with the strategy method. In an ultimatum game with binary offers, changing the framing from *'exactly equal splits'* to *'nearly equal splits'* leads to lower rejection rates with the direct response method, whereas no changes occur with the strategy method (Güth et al. 2001). Rauhut and Winter conclude that "both methods measure different concepts. The strategy method elicits the normative principles of the respondents, while the acceptance decision in the response game the intensity with

which there principles are adhere to" (p. 1191).

As the comparison between the experiment of Falk et al. (2003) using the strategy method and the replication of Cox and Deck (2005) using the direct response method indicates, the two methods can lead to different results. Brandts and Charness (2011) do not support the expectation that the direct response method is 'hot' and, therefore, more emotional. Consequently, emotions may not (fully) explain the differences between the two methods (pp. 387-390). In a simple game, such as the dictator or ultimatum game, where participants have a limited set of choices, the complexity of the decision situation is low. In these studies, the strategy method does not lead to significantly different behavior than the direct response method, suggesting that the participants' behavior is internally consistent (Cason and Mui 1998; Brandts and Charness 2000).

Summary 6.2: Elicitation Procedures

The results presented should be taken with some caution, as the number of studies directly comparing the two methods is limited. However, since finding treatment effects is the primary interest of almost all laboratory experiments, no paper finds a treatment effect using one method and not the other. The strategy method simply lowers the threshold for potential treatment effects (Brandts and Charness 2011, p. 392). In general, the data collection part is more accessible and cheaper with the strategy method. However, it comes at the cost of abstraction from natural settings, which may affect the magnitude of treatment effects but not the occurrence of an effect (Brandts and Charness 2000).

In our experimental setting, it makes sense to use the (vector) strategy method in the first part to measure the normative punishment preferences of the participants and to have complete information about how someone would punish an equal or unequal allocation. Specifically, we want to know if a participant always or never punishes and, if he punishes, the amount of punishment points assigned. In addition, we are interested in comparisons of punishment behavior between the two groups. For example, if someone does not punish in part one, but does so in part two, an analysis of the context in part two may provide meaningful insights into the factors influencing the punishment decision. Alternatively, if a participant punishes both allocations in part one, showing a general tendency to punish regardless of the outcome, it is not surprising that he (always) punishes in part two as well. This may lead to the conclusion that individual characteristics or preferences, rather than

the situational context, determine the punishment decision. However, using the strategy method may lead to a higher level of assigned punishment points than the direct response method, which should be taken into account when analyzing the punishment behavior of the first part.

In the second part of the experiment, the sequential dictator game with punishment, it is necessary to use the direct response method because it would not be possible to implement the timeout option with the strategy method. In addition, as Rauhut and Winter (2010) note, the direct response method elicits the intensity with which normative principles, here punishment preferences, are followed. Even though this means applying a different method than in the original paper by Bartling et al. (2015), it is reasonable and useful to use the direct response method since the possibility to passively implement an allocation is one of the main elements of our study. Since the dictator game is a simple decision situation and the participants' choice set is limited to two options, the decision situation is of low complexity. Consequently, behavioral differences between the two methods are expected to be small. If anything, the use of the direct response method might lead to a lower level of punishment points than the strategy method. In addition, as the previous discussion has shown, no differences in treatment effects are expected when comparing the two methods. If treatment effects are found with one method, they are also likely to be found with the other method.

6.4.2. Replications in Experimental Economics

Recently, there has been a growing interest in replications of experimental findings due to "concerns about inflated findings in empirical and experimental analyses" (Camerer et al. 2016, p. 1433). In this context, the term "*credibility crises*" is often used, since a considerable number of findings are in fact not valid and cannot be confirmed in further studies. Replication of results is important because it usually strengthens their robustness and increases their trustworthiness (Maniadis et al. 2017, p. 209).

However, there are different definitions of what exactly constitutes a replication. According to Levitt and List (2009), a reproduction of the original results with the same protocol using different participants is often referred to as a replication (Maniadis et al. 2017, p. 218). In this regard, Berry et al. define "a replication as any project that reports results that speak directly to the veracity of the main hypothesis of the original paper" (Berry et al. 2017, p. 27). Our study uses the experimental design of the original paper and extends it to two dimensions. Thus, the main result of the original study determines a particular hypothesis in our study. Therefore, according to the definition of Berry et al. (2017), it is reasonable to

speak of a replication.

Typically, two aspects are relevant when talking about replications: first, to examine how many studies have been replicated, as in Berry et al. (2017), and second, to replicate existing findings, as in Camerer et al. (2016), to examine the robustness of the findings.

Berry et al. (2017) examine the replication rate of all empirical papers in a volume of the American Economic Review (AER). They take the hundredth volume in 2010, which contains 70 papers, and check how often they have been replicated six years later. To do this, they collect studies from the Web of Science that cite the original paper. Their final sample of 1546 cited papers is classified along three dimensions: replication, extension, and (or) robustness check. Overall, 60% of the 70 original papers have one or more citations that fall into (at least) one of the three dimensions, with the majority being robustness checks and (or) extensions of the original paper. Only 29% of the papers have at least attempted replication, with highly cited papers having a higher chance of being replicated (Berry et al. 2017, pp. 27-29). In addition, Maniadis et al. (2017) examine replication rates in the top 150 economics journals, with a focus on experimental economics. They identify about 4.2% of experimental studies to be replicated, with almost half of them succeeding in replicating the original results (pp. 224-227).

The "Reproductibility Project: Psychology" (RPP) replicates 100 studies published in the top three psychology journals using the original (provided) material. In the absence of established replication criteria, they use five factors: significance, p-values, effect sizes, subjective ratings, and meta-analyses of effect sizes. Overall, there is only weak support for the original findings in the replications. For example, the number of studies with significant results in the replication sample is 61 percentage points lower than in the original studies (97% to 36%, $p < 0.05$). In addition, only half of the original effect sizes are within the confidence interval (95%) of the replication effect size.

Camerer et al. (2016) attempt to replicate 18 laboratory experiments published between 2011 and 2014 in two top journals, the AER and the Quarterly Journal of Economics (QJE). In order to replicate the most statistically significant results of each of these experimental studies, they use the same replication indicators as in the RPP. More specifically, they focus on three factors: (1) the direction of the significant effect, (2) the 95% confidence interval, and (3) a meta-analysis estimating the effect size (pp. 1433-1434). Overall, 11 out of 18 laboratory experiments can replicate a significant effect in the same direction as in the original study. In 12 out of 18 studies, the confidence interval of the replication includes the original effect size. Furthermore, the meta-analysis finds a significant effect in the same direction as the original study in 14 replications. A comparison with the RPP replication results supports

the robustness of empirical findings from laboratory experiments. As datasets and instructions are usually provided in experimental economics, the high transparency explains these results (Camerer et al. 2016, pp. 1434-1435).

Summary 6.3: Replications in Economics

Replication of experimental and empirical findings is important because it strengthens the findings and serves as a further robustness check. Although it is not always possible to replicate original findings, the successful replication rate for laboratory experiments is high because datasets are usually made available, providing greater transparency.

Our study uses the experimental design of Bartling et al. (2015), but does not replicate it in a narrow sense. Instead, this study extends the original experiment by implementing the treatment variations to examine the robustness of their findings. Previous replications and extensions by Duch et al. (2014) and Anselm et al. (2022) find similar results to the original study. Duch et al. (2014) replicate the original experiment and add a proposer stage. Their replication confirms the main finding of Bartling et al. (2015) that the pivotal decision-maker is seen as most responsible. The implementation of the proposer stage shifts the assigned responsibility to the proposer if the proposed allocations are not fair to the recipients. Anselm et al. (2022) extend the group size to ten (five decision-makers and five recipients) and allow for rewards and punishments. Comparing their punishment treatments to the original study by Bartling et al. (2015) supports also the superior role of the pivotal decision-maker in attributing responsibility.

6.5. Hypotheses

We formulate the following hypotheses about the recipients' punishment behavior in our experiment based on our working hypotheses, the synthesis in Chapter 5, and the pretest. As we have already specified in the previous section, we have to assume that responsibility and punishment are equivalent to each other. More precisely, holding someone responsible automatically leads to punishing him to the extent of the responsibility assigned. Consequently, the following hypotheses are formulated regarding the punishment decision.

This section is divided into three parts. First, we focus on punishment motives based on the Working Hypotheses 1, 2, 6, 7, and 8. Second, we consider our treatment variation. More specifically, we specify punishment expectations regarding the default (Working Hypothesis

3) and legitimate status differences (Working Hypotheses 10, 11, and 12) as well as the interaction between both dimensions. Third, individual characteristics (Working Hypotheses 5 and 9) and expectations (Working Hypothesis 4) are taken into account and related hypotheses are defined.

6.5.1. Punishment Motives

The use of punishment is one way of sanctioning norm violations. Relatedly, altruistic punishment implies the costly use of punishment without any material reward or gain from it (Balafoutas et al. 2014, p. 15924), as is the case in our study. However, a purely self-interested recipient would never punish because it costs him a point and brings him no material benefit. Although empirical evidence has shown that people do not always behave as predicted by the theory (among others, see e.g. Fehr and Gintis 2007, p. 45; Akbaş et al. 2019), there may be recipients who act in a self-interested, payoff-maximizing way. This is captured by our first hypothesis:

Hypothesis 1 (Rationality) *A rational and self-interested, payoff-maximizing recipient never punishes.*

However, experimental and empirical evidence suggests that people are not as self-interested and rational as the theory predicts, since the payoff of others is also part of their utility function (Fehr and Gintis 2007, p. 45). Even in one-shot interactions, where people never interact again, costly punishment often occurs (for a review, see Chaudhuri 2010). Since punishment in this study is costly and provides no material benefit, the punishment behavior of recipients can be explained by social preferences and the decision context (Akbaş et al. 2019), which we capture with the following hypotheses.

As stated in Working Hypothesis 6, people are inequality averse (Fehr and Schmidt 1999; Bolton and Ockenfels 2000) and consider only the implemented outcome. In this context, responsibility is assigned according to the outcome and not according to individual choices. Consequently, we expect no punishment if the outcome is equal, while all decision-makers are punished equally if the outcome is unequal.

Hypothesis 2 (Outcome) *When the unequal allocation is implemented, all decision-makers are punished equally, while no punishment is assigned for an equal outcome.*

While the motive 'outcome' focuses solely on the implemented allocation, the motive 'intention' addresses the decision-maker's individual choices. This motive has been specified

in Working Hypothesis 7 and is related to intention-based models (Rabin 1993; Dufwenberg and Kirchsteiger 2004). Typically, the prediction of punishment depends on the beliefs of the participants. However, even without knowing these beliefs, as long as the outcome is not fixed, choosing the unequal allocation always leads to a higher probability that the unequal outcome will be implemented. Therefore, choosing the unequal allocation is seen as unkind and is punished, regardless of the outcome that is actually implemented. Even if one decision-maker chooses unequal and the other two choose equal, the decision-maker who chooses unequal is expected to be punished. Thus, Hypothesis 3 specifies that any decision-maker who chooses the unequal allocation is punished more than any decision-maker who chooses the equal allocation.

Hypothesis 3 (Choice) *Choosing the unequal allocation leads to more punishment.*

Furthermore, even more punishment is expected if the choice is impactful, as defined in Hypothesis 4. In addition, regarding Working Hypothesis 8 and the combination of outcome- and intention-based models (Falk and Fischbacher 2006; Falk et al. 2008), it can be expected that the assigned punishment is even higher if the implemented outcome is unequal.

Hypothesis 4 (Unkindness) *Punishment points are awarded as long as the choice of the unequal allocation has a bearing on the outcome.*

The Working Hypothesis 1 focuses on the first decision-maker in a decision sequence because he initiates the sequence. This assumption is based on the theoretical work of Spellman (1997) and tested experimentally by Duch et al. (2014). Applying this to our design, we expect that the first decision-maker, the initiator, who chooses the unequal allocation is punished more than the subsequent decision-maker(s), especially when the implemented outcome is unequal. Therefore, Hypothesis 5 is defined as follows:

Hypothesis 5 (Initiation) *The first decision-maker who opts for the unequal allocation receives more punishment points than the following decision-maker(s).*

The next punishment motive, pivotality, is based on Working Hypothesis 2 and states that the pivotal decision-maker is punished more than the non-pivotal decision-makers. Thus, the second decision-maker who chooses unequal, and therefore sets unequal as the final allocation, is punished more than the other decision-makers (independent of their choices). Hypothesis 6 refers to the experiment by Bartling et al. (2015) and Anselm et al. (2022), as they find that the pivotal decision-maker receives significantly more punishment points than the non-pivotal decision-maker. In addition to the experimental confirmation, this is also consistent

with the theoretical framework of the Structural Model Approach of Chockler and Halpern (2004) and Engl (2018).

Hypothesis 6 (Pivotality) *The pivotal decision-maker receives more punishment points than the non-pivotal decision-maker.*

Obviously, some Hypotheses overlap in some ways, or rather contradict each other. For example, a pivotal decision-maker can also be classified as having an unkind intention. In addition, either Hypothesis 5 or Hypothesis 6 is applicable, because either the first (= initiator) or the second (= pivotal) decision-maker who opts for unequally is punished more. The experimental results of Bartling et al. (2015) and Anselm et al. (2022) suggest that Hypothesis 6 is more likely than Hypothesis 5. However, the role of initiation should not be overlooked.

6.5.2. Treatment Variations

Apart from the punishment motives, the two variations, default and mechanism, and their interaction are expected to influence the assigned punishment points. In addition, we expect them to influence the strength of the previously presented punishment motives.

Many experimental studies show that acts of omission are less punished than acts of commission (among others, see e.g. DeScioli et al. 2011b; Hayashi 2013). Accordingly, as stated in Working Hypothesis 3, choosing the default option is expected to be less punished because it can be passively implemented by an act of omission. Consequently, Hypothesis 7 claims that the decision-maker who chooses the default option is punished less. In particular, when the unequal allocation is preselected, less punishment is expected for an unequal choice than for an unequal choice with an equal default. Recipients cannot unambiguously infer the underlying intention of the decision-maker (Vaal 1996, see e.g.), as the default can be implemented actively (by clicking the continue button) or passively by doing nothing. To combine this, Hypothesis 7 is twofold, covering the different (potential) leverage of a default.

Hypothesis 7 (Default) *(a) More punishment is expected when the equal allocation is preselected because the unequal allocation must be actively implemented. (b) Choosing the default leads to less punishment.*

People with higher status, as we define decision-makers in our experiment, are seen as competent and therefore more deserving and entitled to rewards (Working Hypothesis 11). Even if the decisions are unfair and reinforce inequalities, they are more likely to be accepted (Albrecht et al. 2013; Fleiß 2015). In addition, consistent with Working Hypothesis

10, legitimate status differences induced by fair procedures are expected to lead to higher acceptance of unequal outcomes, as people may perceive them as less unfair (Hegtvedt et al. 2003, pp. 355-356). Putting both aspects together, as clarified in Working Hypothesis 12, we assume that unequal outcomes are more likely to be accepted when status differences are legitimate, so fewer punishment points are imposed on decision-makers with a legitimate decision right. Consequently, the more legitimate a random procedure is perceived to be, the more randomly induced status differences should be accepted, leading to less punishment for (unequal) decisions by high-status participants.

Hypothesis 8 (Status) *A more legitimate status structure leads to a higher acceptance of inequalities, so unequal choices of randomly assigned decision-makers are less punished.*

Turning to the interaction effect of default and status differences is not so straightforward. Consequently, predictions are not easy to make since no previous study has combined the two. However, it is expected that the combination of an unequal default and a random group-building mechanism result in the fewest punishment points. On the other hand, we expect to see the most punishment points in grade-equal, because the group-building mechanism is less legitimate and the unequal allocation has to be actively chosen.

Specifying the two extreme positions can be done based on the previously stated Hypotheses 7 and 8. However, it is not predictable whether the default option or the group-building mechanism has a higher impact on the punishment decision. If the default had a larger positive impact, one would expect more punishment in random-equal than in grade-unequal. Conversely, if the group-building mechanism is a stronger driver of the punishment decision, more punishment is likely in grade-unequal than in random-equal. Based on the experiment by Duch et al. (Duch et al. 2014), where a proposer is held more responsible than participant for not using his veto, we expect the latter, i.e., a larger effect of the group-building mechanism than of the default.

Hypothesis 9 (Interaction) *We expect the least punishment in random-unequal, because status differences are legitimate and an unequal allocation can be implemented passively. In contrast, a less legitimate mechanism and the implementation of an unequal allocation by an act of commission (grade-equal) are hypothesized to be punished the most.*

6.5.3. Individual Characteristics and Beliefs

In this section, we begin with individual characteristics that may influence punishment decisions. Although empirical and experimental evidence is scarce and sometimes unambitious,

women are more inequality averse (Andreoni and Miller 2002; Dufwenberg and Muren 2006), more loss averse (Schmidt and Traub 2002) and more risk averse (Holt and Laury 2002; Croson and Gneezy 2009) than men, which is expected to affect their punishment decision. Moreover, in a public goods game, women are slightly more likely to punish (Burnham 2018), while men are more likely to support punishing unfair behavior (Singer et al. 2006). Consequently, as already specified in Working Hypothesis 5, we can specify our next hypothesis about gender differences in punishment.

Hypothesis 10 (Gender) *We expect women to punish unfair behavior less than men. This concerns the probability of punishment and the amount of punishment points assigned.*

The influence of risk attitudes on punishment behavior is still an open question, so we can only speculate about the influence in our experiment. However, as we stated in Working Hypothesis 9, we can assume the following relationship:

Hypothesis 11 (Risk Attitude) *We expect risk-averse participants to punish less than risk-seeking participants.*

Norms define appropriate behavior in specific situations and thus influence individual behavior (Elster 1989; Bernhard et al. 2006). Violating a norm is likely to be sanctioned, often by inducing monetary punishment, which has been demonstrated in a wide range of experimental and empirical findings (among others, see e.g. Chaudhuri 2010; Balafoutas et al. 2014). In addition, evaluating the intention of decision-makers depends on the recipient's beliefs about their motives for choosing an allocation (Dufwenberg and Kirchsteiger 2004). As proposed by Çelen et al. (2017), the evaluated kindness of a decision depends on whether the recipient would act in the same way. Combining both aspects with our assumption in Working Hypothesis 4, we specify the following hypothesis.

Hypothesis 12 (Belief) *If a recipient chooses the same allocation as the decision-maker, we expect him to allocate fewer punishment points.*

Table 6.4 gives an overview of our hypotheses and their corresponding punishment predictions, which are analyzed in the next section.

Table 6.4.: Punishment Predictions

		Hypothesis	Punishment Prediction
Punishment Motives	H1	Rationality	A rational recipient does not assign punishment points, regardless of the choice and outcome.
	H2	Outcome	All decision-makers are punished equally if the unequal allocation results.
	H3	Choice	A decision-maker choosing the unequal allocation is punished more.
	H4	Unkind	Decision-makers who choose unequally and are impactful are punished more.
	H5	Initiation	The first decision-maker who chooses unequal is punished the most.
	H6	Pivotality	The pivotal decision-maker is punished the most.
Treatment Variations	H7	Default	(a) An active implementation of the unequal outcome is punished more. (b) Not choosing the default is punished more than choosing the default.
	H8	Status	When the final grade determines status differences, unequal choices are punished more severely.
	H9	Interaction	We expect the least punishment in random-unequal, while a less legitimate mechanism and the implementation of an unequal allocation by an act of commission (grade-equal) are hypothesized to be punished the most.
Individual Characteristics	H10	Gender	It is expected that unfair behavior is punished more often and more severely by men than by women.
	H11	Risk	We expect risk-averse participants to punish less than risk-seeking participants.
	H12	Belief	If the recipient makes the same decision as the decision-maker, we expect less punishment.

7. Results

The chapter is divided into four sections. In Section 7.1, we begin by describing our sample and the punishment behavior from the first part of the experiment. To ensure that participants are from the same subject pool, we test for potential heterogeneity across treatments. Although we are not primarily interested in analyzing the behavior of the decision-maker, in Section 7.2 we provide a brief overview of how they decide and what decisions the recipients expect. Section 7.3 provides a rating comparison of the group-building mechanism between the pretest and the experiment, and also covers role deservingness.

Section 7.4 is the main part of this chapter and examines how participants actually punish. Therefore, we present a classification of decision-makers based on the hypotheses identified earlier. After that, Section 7.4.2 first presents an overview of punishment behavior, focusing on the two distinct motives of 'initiation' and 'pivotality'. Two different models are estimated to examine how participants assign responsibility. First, a probit model determines the factors that influence the decision to punish at the extensive margin, i.e., how many people punish regardless of the amount of assigned punishment points (Section 7.4.3). Second, a tobit regression estimates the treatment effect at the intensive margin, i.e., the (exact) amount of punishment points assigned (Section 7.4.4). Thus, the first analysis attempts to disentangle whether a recipient punishes, while the second analysis examines the amount of punishment assigned. We also include the BSJO scale, various robustness checks, and present the reported answers to the open question of how participants decided. In addition, we estimate separate analyses for the default and the mechanism, as well as a hurdle model and a cluster analysis as further controls to strengthen our results (Section 7.4.5). Finally, in Section 7.4.6, we evaluate a comparison between the results of Bartling et al. (2015) and our results, considering our aggregated dataset without treatment differences. The last section of all, Section 7.4.7, discusses the expected punishment of the decision-makers, i.e., how they perceive their own responsibility.

7.1. Breakdown of the Sample

7.1.1. Socio-demographic Characteristics

Table 7.1 summarizes our sample according to the data collected in the questionnaire at the end of the experiment. Because of the requirement to answer every question, there are no missing values in our dataset. Most participants are female (176 out of 270, 65.19%), with a similar distribution across treatments and sessions. Only one session (session 8, treatment 1: random-equal) achieves an almost equal gender distribution. Using a χ^2-test, no significant differences ($p = 0.763$) are found between the treatments. On average, the participants are 26.77 years old and in their 8.4th semester. 40.37% study (socio-)economics or law, 25.56% study natural sciences and 27.04% study humanities. These patterns are similar across treatments. Consequently, there are no significant differences between the treatments in terms of these socio-demographic characteristics, so our sample is homogeneous and comparisons are possible.

Table 7.1.: Breakdown of the Sample

		N	Female	Age	Semester	Mean grade
Random-unequal	(T1)	72	66.67%	26.23	6.71	2.02
Grade-unequal	(T2)	66	62.12%	26.86	9.12	2.00
Random-equal	(T3)	66	62.50%	26.99	8.69	2.08
Grade-equal	(T4)	66	69.70%	27.00	9.03	2.12
Overall		270	65.19%	26.77	8.40	2.06

Finally, we want to look at the final grade, since this induces the status differences. As we can see from Table 7.1, the average grade is similar for all the treatments. Furthermore, there are no significant differences between random and grade treatments (T1/T2 vs. T3/T4, two-tailed t-test, $p = 0.857$), nor between grade treatments (T3 vs. T4, two-tailed t-test, $p = 0.255$). However, there are significant differences between males and females. On average, women have a better grade than men (2.00 vs. 2.16), which is significant at the 5%-level. This difference is significant only in the random treatments and not in the grade treatments (2.03 vs. 2.12, two-tailed t-test, $p = 0.406$), indicating that the effect on group formation should be minimal. In a cross-national meta-analysis of gender differences in school achievement, Voyer and Voyer (2014) confirm that women get better grades in most courses in school,

although they do not do better on achievement tests where men do better in mathematics and sciences. Women's superior academic performance is present at all levels of schooling: elementary, middle, high school, and college (p. 1174).

It can be argued that participants may lie about their grade and report a better (worse) grade than they actually receive. Participants who are honest may assume that others have not been as honest as they have been. As a result, the credibility of the final grade would suffer. There is a large experimental literature on lying and cheating (among others, see e.g. Abeler et al. 2019; Fischbacher and Föllmi-Heusi 2013). For example, in the famous roll-a-dice experiment by Fischbacher and Föllmi-Heusi (2013), nearly 40% of participants do not lie at all, 20% of inexperienced participants lie as much as possible, while the rest are declared as partially lying in the sense that they lie but not to the maximum extent. Thus, we would also expect participants to lie in our study by reporting a better (worse) grade than they actually received. However, in contrast to previous studies, participants in our design do not have an incentive to be dishonest. Rather, they do not know at this stage what the final grade will be used for and whether it might be advantageous to report a better (worse) grade than they actually received. Moreover, the average reported grade in our experiment is 2.06, which is higher than the average reported final grade in the 2020/2021 school year (see Section 6.1.4). Given that not all graduates go on to university and many programs require a specific grade, this could be due to sample selection bias rather than lying.[53] However, we cannot completely rule out lying in our sample and should be aware of the (potential) negative effect on the credibility of the final grade.

Unfortunately, we find significant differences between participants in their self-reported risk attitudes. On a five-point Likert scale ranging from 1 (strongly risk-taking) to 5 (strongly risk-averse), participants indicate where they place themselves. Table 7.2 shows the proportion of risk-averse individuals in our sample, separated by gender and treatment. As expected, and as has been demonstrated in a wide range of empirical findings (among others, see e.g. Holt and Laury 2002; Fehr-Duda et al. 2006; Croson and Gneezy 2009), women in our sample are more risk averse than men (44.31% vs. 21.27%). A χ^2-test reveals statistically significant differences in self-reported risk attitudes between men and women ($p < 0.001$), which was already expected. However, we also have significant differences in self-reported risk attitudes

[53] Since we ran the experiment at the University of Hamburg, we can take a look at the bachelor's entry degrees (NC) for the most represented programs identifiable in our sample over the last six years. Although some programs do not have a minimum degree requirement, there are many programs that do. In business administration, which 17% of our sample study, the NC lies between 2.2 and 2.6, in psychology (6% of the sample) between 1.2 and 1.4, in law (6% of the sample) between 1.9 and 2.4 and in economics (5%) between 2.7 and 3.1 with being open to everyone in 22/23 (Hamburg 2022).

Table 7.2.: Self-Stated Risk Attitude

		N	Risk affine	Risk neutral	Risk averse	χ^2-test
Overall		270	30.37%	33.33%	36.29%	
Male		94	42.55%	36.17%	21.27%	$\chi^2 = 99.43 \quad p < 0.001$
Female		176	23.86%	31.82%	44.31%	
Random-unequal	(T1)	72	31.94%	33.33%	34.72%	
Grade-unequal	(T2)	66	33.33%	27.27%	39.39%	$\chi^2 = 87.17 \quad p < 0.001$
Random-equal	(T3)	66	25.76%	33.33%	40.91%	
Grade-equal	(T4)	66	30.30%	39.39%	30.31%	

Table notes: Risk-attitude is self-stated on a 5-point Likert scale. Risk affine correspond to strongly risk affine and riskaffine. Risk averse correspond to strongly risk averse and risk averse. χ^2-tests are calculated using the 5-point Likert scale.

between treatments ($\chi^2 = 87.17, p < 0.001$), which is unexpected and suboptimal, as punishment differences between treatments may be driven by differences in risk attitudes. It is also possible that there are no punishment differences between treatments because participants differ in their risk attitudes, which may obscure the differences between treatments. Alternatively, these differences may be driven by the unequal proportion of women in the four treatments. Either way, we will return to this point when we present our regression analysis.

Of course, it is not ideal for females to have a better grade in our sample, as this could lead to more males being given the role of 'recipient' in the grade treatments. However, controlling for gender and risk attitude in the regression analysis should eliminate or mitigate any potential effect. Nevertheless, the validity of our sample is ensured to the greatest extent possible by the fact that participants are drawn from the same subject pool and differ only slightly between treatments.

7.1.2. Homogeneity of the Sample

This section presents the results of the first part, the 1:1 dictator game with punishment. As a reminder, the experiment starts with a two-person dictator game in which each participant acts as a dictator and decides on either an equal (5;5) or an unequal (9;1) allocation. Then, using the strategy method, each participant can assign up to five punishment points for both distributions. Deducting points costs one point, no matter how many points are deducted.

Table 7.3.: Part 1 - Dictator Game with Punishment

		Dictator		Recipient			
				EQUAL		UNEQUAL	
	N	Equal [in %]	Unequal [in %]	Punishment [in %]	Mean	Punishment [in %]	Mean
Overall	270	60.37	39.63	11.11	0.31	68.15	3.20
Male	94	58.51	41.49	14.89	0.47	70.21	3.30
Female	176	61.36	38.64	9.09	0.23	67.05	3.18
Random-unequal	72	63.89	36.11	15.28	0.44	76.39	3.51
Grade-unequal	66	60.61	39.39	6.06	0.26	68.18	3.36
Random-equal	66	62.12	37.88	15.15	0.35	66.67	3.06
Grade-equal	66	54.55	45.45	7.58	0.20	60.61	2.83

Table notes: The equal allocation distributes 5 points to the dictator and recipient. The unequal allocation distributes 9 points to the dictator and 1 point to the recipient. For each allocation, the assignment of up to 5 punishment points is possible. Punishment in % is the proportion of participants who punish. The mean indicates the average assigned punishment points for the allocation including those who do not punish.

This part of the experiment aims to elicit participants' general punishment behavior when faced with an unequal or equal allocation. In addition, because this data is collected prior to the group-building stage, it serves as a verification of similar punishment norms within participants and across experimental parts.

Table 7.3 displays the choices as dictators and the punishment points awarded as recipients. A quick look at the results shows that 39.63% choose the unequal offer as dictator. Males are slightly more selfish (41.49% unequal choices), which is not statistically different from females (two-tailed t-test, $p = 0.649$). A χ^2-test finds no statistically significant differences in the distribution of choices between treatments.

Next, we turn to the punishment decision as recipient. As a reminder, rational choice theory predicts no punishment because punishment is costly and provides no (material) benefit. However, contrary to the theoretical prediction, participants allocate punishment points. On average, choosing the equal distribution is less punitive than choosing the unequal distribution (0.31 points vs. 3.20 points). Almost 90% of participants do not punish an equal choice. This is consistent with both rationality and inequality aversion, as well as a general preference for equal payoffs. However, when participants do punish, they either choose the

maximum amount of five points (which is perfectly logical since recipients have to pay one point either way, so why not assign the maximum amount) or only one point. Interestingly, men punish an equal allocation slightly more than women do (0.47 points vs. 0.23 points), which is barely significant at the 10%-level (two-tailed t-test). This punishment behavior cannot be explained by the theories and models of social preferences presented so far. If anything, the participants can be classified as spiteful or envious, since they destroy an equal division and lose a point in order to make the other person worse off than themselves. There is a large literature on antisocial punishment confirming that there are people who punish others for prosocial or cooperative behavior (among others, see e.g. Levine 1998; Herrmann et al. 2008).

Again, not surprisingly, participants assign more punishment points for choosing the unequal allocation. However, 31.85% do not punish, as predicted by rational choice theory. Most participants (58.52%) who punish assign the maximum number of punishment points possible (again, not surprising). On average, men attribute more punishment points to an unequal distribution than women do (3.30 vs. 3.18 points), which is not significantly different (two-tailed t-test, $p = 0.612$). When we turn to the treatment distinction, we see that participants in the random-unequal and grade-unequal treatments punish more often (76.39% and 68.19%, respectively, while only 66.67% and 60.61% punish in the other two treatments) and assign more punishment points on average. However, a χ^2-test between the two default conditions reveals no statistically significant differences in the distribution of punishment points between the equal and unequal default groups (neither for the two allocations separately nor for the aggregated data).

To further investigate whether participants differ in their punishment behavior between treatments, a tobit regression is estimated with the assigned punishment points as the dependent variable. Thus, the purpose is to show the homogeneity of the sample in terms of punishment behavior across the treatments. In order to have only one punishment variable, the dataset is reshaped with the punishment decision as the round identifier and declared as panel data. First, default and mechanism are regressed separately on the assigned punishment points. Then the treatment variable is regressed on the punishment variable, before adding several control variables (age, gender, and the offer chosen as dictator). Table A.2 in Appendix A.3 displays the results of these regressions.

Overall, there are no significant differences between treatments. Controlling for default and mechanism reveals that participants in the default treatments punish slightly less (Coef. -0.264, $p = 0.093$). Thus, participants assigned to the unequal default group allocate fewer punishment points in the dictator game *before the group-building phase*. Apart from this

unpleasant effect, the results are as expected or hoped for. One's choice as dictator has a significant effect on the punishment points assigned, and no gender differences are apparent.

Summary 7.1: Homogeneity of the Sample

In summary, the analysis indicates that participants differ only slightly in their punishment behavior between the two default classifications. Since we have only a small sample of a population, this could occur. However, participants do not differ per se between treatments, so our sample can be considered homogeneous. The validity of the data set is confirmed and comparisons between treatments are possible.

7.2. Decision-Makers' Actual Choices

Although our primary interest is not in analyzing the behavior of the decision-makers, it is worthwhile to provide a brief overview of how they decide, which is offered in the first part of this section. In addition, the expected choices of the recipients, i.e., how they would expect the decision-makers to choose, are presented. In particular, the latter may provide insights into recipients' punishment behavior, as their choices may reveal normative standards and beliefs about how decision-makers should behave, thereby influencing their punishment decision.

Of the 135 participants classified as decision-makers, one-third decide as first, second, and third decision-makers in each round. Thus, in each round we have 45 observations per position (A1, A2, A3), resulting in 810 individual and 270 group observations.[54] Of the 270 group observations, 54 (20%) lead to an unequal outcome, while decision-makers chose the unequal allocation 175 times out of 810 (21.6%).[55]

Interestingly, there are significant differences between the default option and the choices, but not between choices and mechanism, as presented in Table 7.4. When the unequal allocation is preselected, participants are more likely to choose the equal allocation than when the equal allocation is declared the default (81.40% versus 75.25%). Similarly, when the unequal allocation is preselected, the unequal outcome occurs 13.77% of the time, while an equal default leads to an unequal outcome 26.52% of the time. Consequently, the unequal outcome is realized more often when the default is equal. In other words, an unequal

[54] In contrast, each recipient decides three times in each round whom and how much to punish.

[55] Interestingly, this is quite different from the results of Bartling et al. (2015), where almost two-thirds of all decisions lead to the unequal outcome.

Table 7.4.: Choices Depending on the Default and Mechanism

		Choices		
		Unequal	Equal	χ^2
Default	Unequal	18.60%	81.40%	$p < 0.01$
	Equal	24.75%	75.25%	
Mechanism	Random	21.26%	78.74%	$p = 0.546$
	Grade	21.97%	78.03%	

default results in an equal outcome more often than in an unequal outcome (a difference of 12.75 percentage points). A default can be perceived as the recommended option or as more acceptable to choose (e.g. McKenzie et al. 2006), so one might expect that the unequal default would lead to more unequal choices, though this is not the case in our experiment.

Performing a χ^2-test on choices and default yields a high χ^2-value ($\chi^2 = 27.11$. The p-value of the test is $p < 0.01$, indicating that the participants' choices differ significantly between the default treatments. Based on this, it can be concluded that participants are more likely to make unequal choices when the default is equal, which leads to more unequal outcomes when the default is equal. Consequently, an unequal choice is more likely to be implemented actively than passively, which is completely unexpected and not as predicted by previous empirical and experimental findings (among others, see e.g. Spranca et al. 1991; Jachimowicz et al. 2019). Why participants in our experiment do not behave as predicted can only be speculated. Perhaps the presence of an unequal default option made them think more about the appropriate (or expected) choice. When we turn to the choices depending on the group-building mechanism, a slightly different picture emerges, since the decision-makers' choices do not differ between the two mechanisms ($p = 0.546$).

A more detailed look at the decision-makers' choices and a comparison with the experiment by Bartling et al. (2015) is worth noting. In Table 7.5, the last three columns show the decision-makers' choices, the recipients' expectations of how the decision-makers would choose, and the decision-makers' choices in the Bartling et al. (2015) experiment. In our study, the unequal allocation is implemented in 20.0% of the rounds, and would have been implemented similarly often according to the recipient's expectations. Turning to the individual expectations at each decision node, some interesting patterns emerge, especially when compared to Bartling et al. (2015).

Table 7.5.: Choices and Expected Choices at each Decision-Node

Decision-maker	Sequence (decision in bold)	This study		Bartling et al. (2015)
		Decision-maker	Recipient	
1	**u-...**	22.59	26.30	58.3
	e-...	77.41	73.70	41.7
2	u-**u...**	39.34	38.03	59.7
	u-**e...**	60.66	61.97	40.3
	e-**u...**	20.10	14.07	61.1
	e-**e...**	79.90	85.93	38.9
3	u-u-**u**	20.83	66.67	22.2
	u-u-**e**	79.17	33.33	77.8
	u-e-**u**	40.54	38.64	70.8
	u-e-**e**	59.46	61.36	29.2
	e-u-**u**	35.71	35.71	62.5
	e-u-**e**	64.29	64.29	37.5
	e-e-**u**	7.78	6.43	1.4
	e-e-**e**	92.22	93.57	98.6
Unequal outcome		20.00	20.00	67.4

Table notes: "u" denotes a choice of the unequal allocation; "e" stands for a choice of the equal allocation.

The first two rows show how often the first decision-maker chooses equal and unequal. The first decision-maker's choice is of particular interest because it is the one who starts the decision sequence. Consistent with the high number of equal choices overall, the first decision-maker chooses the unequal allocation in 22.6% of all rounds, 26.3% of recipients expect the unequal choice, while more than 58.3% of decision-makers choose unequal in Bartling et al. (2015). Turning to the second decision-maker, the picture is quite different. Here, 39.3% and 20.1% of the decision-makers choose unequal, which is again quite close to the expectations of the recipients (38.0% and 14.1%), while in the experiment of Bartling et al. (2015) about 60% choose unequal. If the first decision-maker chooses unequally, the second decision-maker can either determine the outcome by choosing unequally or postpone the final decision to the third decision-maker. As discussed in the theoretical section, postponing or avoiding decisions is often used to avoid punishment or to maintain a positive self-image. In this study, this is the case in 60.7% of the decisions (Sequence u-e), while in Bartling et al.

(2015), only 40.3% avoid the final decision as a second decision-maker.

Furthermore, the choice of the third decision-maker differs between this study and Bartling et al. (2015). Suppose the third decision-maker is pivotal and can therefore determine the outcome because the sequence is either u-e or e-u. In this case, the unequal outcome is implemented in 40.5% and 35.7% in our study and 70.8% and 62.5% in the study by Bartling et al. (2015).

Summarized, the similarities between actual and expected choices strengthen our results, especially in the delineation with Bartling et al. (2015), where a different behavior is observed, which is further analyzed in Section 7.4.6. These differences may be due (in part) to methodological differences, as Bartling et al. (2015) use the strategy method, whereas we apply the direct response method.

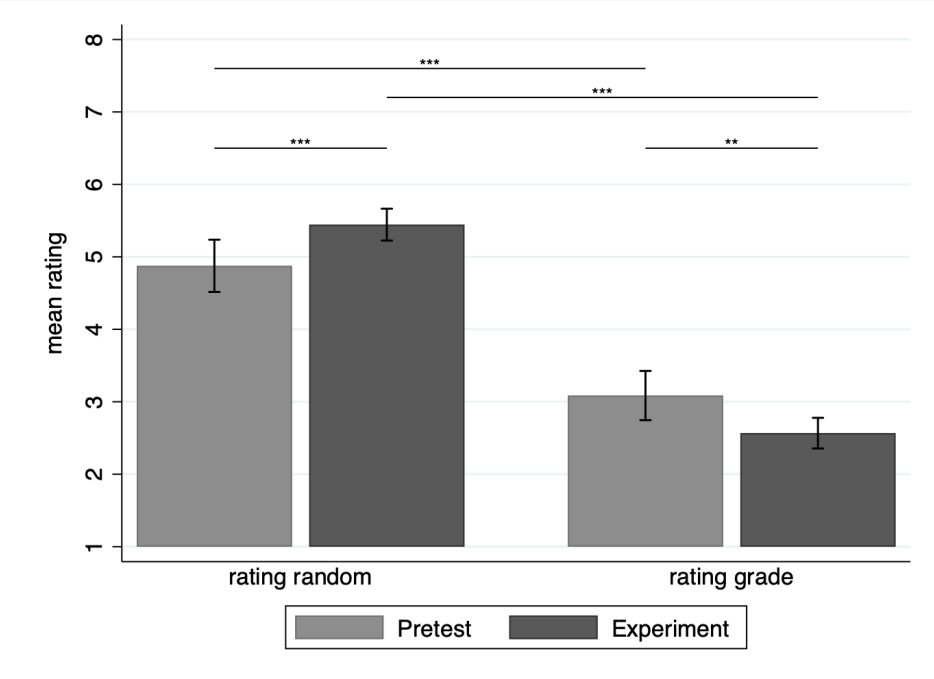
An interesting result in our experiment is that almost no one lets the timer run out or uses the 'do nothing' option to passively implement an unequal distribution. In total, in 2 out of 810 decisions, the 'next button' is not clicked, so the timer runs out. In these cases, however, the equal allocation is preselected, so the participants (probably) simply forget to actively decide in time. Of course, the recipients do not know whether the choice is active or passive. Since the time-out option has no effect on decision-makers' behavior and too few people use it, it is not analyzed further here.

7.3. Rating and Deservingness

Before turning to the estimation of responsibility attribution as the core of our analysis, a brief overview of the rating and deservingness is warranted. To compare the pretest and the experiment, we also assess the evaluation of the group-building mechanism at the end of the experiment. First, participants are asked about the legitimacy of their group-building mechanism. Then they evaluate the other mechanism, even though it is not used in their session. As in the pretest, the same 7-point Likert scale is utilized. In Section 7.3.1 the results of this experiment and a comparison to the pretest are reported. Following Demiral and Mollerstrom (2018), the deservingness of one's own role (either as a decision-maker or as a recipient) and the other role is assessed on a 10-point Likert scale. These results and a comparison with Demiral and Mollerstrom (2018) are discussed in Section 7.3.2.

7.3.1. Rating of Group-Building Mechanism

Figure 7.1 depicts the mean rating of the two mechanisms in the pretest and in the experiment. Since the participants in the experiment are in fact placed in different roles using one of the two mechanisms, it is interesting to see if the rating is different in the experiment compared to the pretest.



Notes: Rating scale range from 1 (completely not legitimate) to 7 (completely legitimate). The p-values are from two-tailed t-tests of treatment differences, assuming equal variances. Bars represent the 95% confidence interval. $^{*}p \leq 0.1$, $^{**}p \leq 0.05$,$^{***}p \leq 0.01$.

Figure 7.1.: Rating of Group-Building Mechanisms - Pretest versus Experiment

In the experiment, random is perceived as much more legitimate than grade. The mean difference between the two mechanisms is 2.88 points (statistically significant at the 1%-level), which is even higher than in the pretest, where it was 1.79 points (same significance level). As it can be seen in Figure 7.1, the rating of the two mechanisms differs slightly between the pretest and the experiment in that participants in the experiment rate a random procedure as more legitimate than participants in the pretest (difference: 0.57 points, two-tailed t-test, $p < 0.01$). Furthermore, the final grade is perceived as less legitimate in the experiment than in the pretest (difference: 0.52 points, two-tailed t-test, $p < 0.05$). Thus, randomly assigned positions are perceived as more legitimate than those based on characteristics acquired in

the past, such as the final grade.

Table 7.6.: Rating of Group-Building Mechanisms - by Role and Mechanism

		Rating random		Rating grade	
		Mean	t-test	Mean	t-test
Gender	Male	4.93	$p = 0.001$	2.84	$p = 0.063$
	Female	5.72		2.42	
Mechanism	Random (T1+T2)	5.09	$p = 0.001$	2.46	$p = 0.295$
	Grade (T3+T4)	5.82		2.68	
Assigned role	*Mechanism*				
Decision-maker	Random	5.01	$p = 0.027$	2.55	$p = 0.340$
	Grade	5.71		2.85	
Recipient	Random	5.16	$p = 0.015$	2.36	$p = 0.608$
	Grade	5.92		2.52	

Table notes: The rating scale range from 1 (completely not legitimate) to 7 (completely legitimate). The p-values are from two-tailed t-tests of treatment differences, assuming equal variances.

Since the participants in the main experiment are actually divided based on one of the two mechanisms, it is reasonable to analyze how a random procedure and the final grade are evaluated when they are used as group-building mechanisms. Furthermore, it is interesting to see if this evaluation differs depending on the role assigned (decision-maker or recipient). Table 7.6 displays the mean ratings by assigned role and mechanism.

As expected, but still interesting, women rate random as more legitimate than men do (difference: 0.79 points, two-tailed t-test, $p < 0.01$), but rate the final grade as less legitimate than men do (difference: 0.42 points, two-tailed t-test, $p < 0.01$), which confirms the pretest. A closer look at the evaluation behavior reveals interesting results. Participants assigned to their role by the less legitimate mechanism (final grade) rate a random procedure as more legitimate than do participants assigned by a random procedure (5.82 vs. 5.09, two-tailed t-test, $p < 0.01$). Combining the two possible roles, decision-maker or recipient, with the two group-building mechanisms, random or final grade, a more nuanced picture emerges. Decision-makers randomly assigned to their role evaluate random as less legitimate than do decision-makers assigned by the final grade (difference: 0.7 points, two-tailed t-test, $p < 0.05$). The same is true for recipients (difference: 0.76 points, two-tailed t-test, $p < 0.05$).

Table 7.7.: Deservingness of Role

		Deserving role A		Deserving role B	
		Mean	t-test	Mean	t-test
Gender	Male	5.72	$p = 0.996$	4.77	$p = 0.255$
	Female	5.72		5.16	
Role	Decision-maker	5.96	$p = 0.169$	4.71	$p = 0.058$
	Recipient	5.48		5.33	
Assigned role	*Mechanism*				
Decision-maker	Random	5.71	$p = 0.326$	4.93	$p = 0.310$
	Grade	6.23		4.48	
Recipient	Random	5.94	$p = 0.041$	5.54	$p = 0.397$
	Grade	5.00		5.12	

Table notes: The deservingness scale range from 1 (completely not deserving) to 10 (completely deserving). The p-values are from two-tailed t-tests of treatment differences, assuming equal variances.

However, the evaluation of the final grade does not reveal any significant differences between the roles and their assignment.

Summary 7.2: Rating of Group-Building Mechanism

Thus, depending on the group-building mechanism and the role assigned, the evaluation of a random procedure differs significantly, while the evaluation of the grade is similar.

7.3.2. Rating of Deservingness

Table 7.7 reports the rated deservingness of the assigned role.[56] Strikingly, there are no significant differences between men and women in their evaluation of the two roles. Although women rate the recipient's role as slightly more deserving (in absolute terms) than men do, these differences are not significant.

Furthermore, there are no significant differences between decision-makers and recipients

[56]For simplicity, the deservingness is displayed as role A (decision-maker) and role B (recipient) as they were labeled in the experiment.

regarding the deservingness of role A. Only weakly, but present, decision-makers perceive the role of recipients as less deserving than recipients perceive their role (difference: 0.62 points, $p < 0.1$), which is completely unexpected. In addition, decision-makers perceive their role as more deserving than recipients do (highly significant at the 1%-level). Specifically, the decision-makers rate the deservingness of their role with 5.96 points and the role of the recipient with 4.71 points (out of 10 points). That is a difference of more than a point. It is interesting to note that the decision-makers consider themselves worthy of their superior role, but at the same time think that the recipients are not worthy of a subordinate role. This is even more unexpected since the recipients do not show any differences between roles A and B. So far, no distinction has been made between the assignment of decision-makers and recipients to their roles.

Since the effect of the group-building mechanism is of primary interest in this study, a further distinction by group-building mechanism is provided, from which two interesting patterns emerge. First, decision-makers who are assigned by their final grades perceive their role as significantly more deserving than the role of recipients (1%-level). In addition, although not statistically significant, decision-makers perceive their role as more deserving when assigned by the final grade than when assigned at random (0.52 points). In terms of the entitlement effect, this is evident because decision-makers earn their role in the grade treatments. For recipients, the opposite is the case. They perceive their role, as well as the role of the decision-maker, as more deserving when it is randomly assigned. For role A, this is significant at the 5%-level.

In Demiral and Mollerstrom's (2018) experiment, decision-makers (they call them proposers) have a significantly higher sense of deservingness for their role in the entitlement treatment than in the random treatment (p. 346). Unfortunately, they only ask proposers, not recipients, about deservingness. Thus, a direct comparison of all combinations is not possible. However, their results are consistent with those in our study, although the differences here are not statistically significant.

Summary 7.3: Rating of Deservingness

Decision-makers consider themselves worthy of their superior role, but simultaneously consider recipients unworthy of a subordinate role. Additionally, decision-makers who are assigned by their final grades perceive their role as significantly more deserving than the role of recipients, while recipients perceive their role, as well as the decision-maker's role, as more deserving when it is randomly assigned.

7.4. Attribution of Responsibility

Since we have 18 observations per subject (6 rounds with 3 positions) that are not independent, we declare our data as panel data with assigned punishment as the dependent variable. In addition, we adjust for session clusters using cluster-robust standard errors in the OLS regressions and clustered by session in the probit regressions.

7.4.1. Classification of Decision-Makers

Before connecting the hypotheses to the analysis of actual punishment, a brief overview of how predictions translate into different punishment motives is useful. Since equal choices and outcomes are not of primary interest in this study, the focus is placed on unequal choices and outcomes.

(1) Outcome unequal:

If the unequal allocation is implemented, the decision-makers are sorted into this category, regardless of their choice. For example, in sequence two (u-u-e), where the third decision-maker chooses equal, he is classified in this category because the unequal allocation is implemented for the whole group. This motive refers to the group-level, while the individual choices determine the following motives.

(2) Unequal choice:

Whenever a decision-maker chooses unequal, he is placed in this category.

(3) Unkindness:

In contrast to 'unequal choice', a decision-maker is classified as unkind as long as his decision has an impact on the implementation of the unequal outcome. Relatedly, an unequal choice increases the probability of implementing the unequal outcome. Consequently, a decision-maker who chooses unequally is classified as having an unkind intention if no majority is reached.

(4) Equal choice:

All decision-makers who choose the equal allocation are placed in this category. When the unequal allocation is implemented, this category is of particular interest because the predictive power of the inequality aversion model is calculated by considering this category (sequence u-u-e, u-e-u, and e-u-u). Since the inequality aversion model specifies equal punishment for all decision-makers if the unequal allocation results, choosing equal should be punished if the unequal allocation is implemented.

(5) Third unequal:

A decision-maker is placed in this category if the unequal outcome is already implemented (the first and second decision-makers choose unequally) and the third decision-maker also chooses the unequal allocation. This category exists only once, namely in the first decision sequence (u-u-u).

(6) Initiation:

The first decision-maker choosing the unequal allocation is classified as initiator. In the original experiment by Bartling et al. (2015), this decision-maker is classified as having an 'unkind intention'. Since he chooses the unequal allocation when the outcome is not yet determined, this implies an unkind intention. This argument is of course possible. However, since this study explicitly focuses on the distinction between the initiator and the pivotal decision-maker, these expressions are used here. Additionally, any decision-maker who chooses unequally can be considered to have an unkind intention, as long as he is influential.

(7) Pivotality:

The decision-maker who determines the outcome is the pivotal decision-maker. Thus, another decision-maker (either the first or the second) has already chosen the unequal allocation, so the pivotal decision-maker can still influence and determine the outcome in both directions, either to an equal or unequal outcome.

Table 7.8 illustrates how the decision-makers correspond to these categories in each of the eight decision sequences.

A note on the generation or classification of punishment motives is worth noting. Of course, there are many ways to define them. Because they often overlap, some decision-makers are placed in more than one category. Whenever a decision-maker is classified as pivotal, it is obvious that he also fits into the category of unkindness and unequal choice, since he necessarily chose unequal to be pivotal. Furthermore, unkindness can be defined in two ways: either every unequal allocation decision is classified as unkind, or only decisions that affect the outcome are classified as unkind, which is the case in this study. The following probit and tobit regression analyses attempt to decipher the determinants of the punishment decision in more detail. Since our definition of unkindness already includes the initiator and the pivotal decision-maker and is quite close to an unequal choice per se, it is not used as a control variable in the probit and tobit regression analyses. However, compared to Bartling et al. (2015), it is considered to show the separate influence of this motive in contrast to the other motives.

Table 7.8.: Categorization of Decision-makers

Sequence	Decision-maker 1	Decision-maker 2	Decision-maker 3	
u-u-u	(2), (3), (6)	(2), (3), (7)	(2), (5)	(1)
u-u-e	(2), (3), (6)	(2), (3), (7)	(4)	
u-e-u	(2), (3), (6)	(4)	(2), (3), (7)	
e-u-u	(4)	(2), (3), (6)	(2), (3), (7)	
u-e-e	(2), (3), (6)	(4)	(4)	
e-u-e	(4)	(2), (3), (6)	(4)	
e-e-u	(4)	(4)	(2)	
e-e-e	(4)	(4)	(4)	

Notes: The table shows which decision-makers are placed into the seven categories defined above for the different sequences. For example, the top row shows the sequence in which all decision-maker chose the unequal allocation (u-u-u). Thus, all three decision-maker are classified into the category (1) unequal outcome and (2) unequal choice. Additionally, decision-maker 1 and 2 are classified as unkind (category (3)), as their choices are still impactful. The first decision-maker is also the one starting the sequence and thus, categorized as (6) the initiator while the second decision-maker is pivotal for the unequal outcome (7). Furthermore, the third decision-maker falls into category (5) as he chooses unequal although the unequal allocation is already implemented.

7.4.2. Overview of Punishment Behavior

Before turning to a joint regression analysis of different punishment motives, a brief overview of punishment behavior in this study is worthwhile. Therefore, this section begins with an overview of how participants punish in general (without treatment distinction). We then look at the average punishment points assigned across treatments and consider how punishing behavior depends on expected choices.

General Punishment Behavior

135 recipients made three punishment decisions in each of the six rounds, so we have 2430 punishment decisions to analyze. In total, 15.85% of the time punishment points are awarded. 42.22% of the recipients never punish, and only one participant punishes in each round. This suggests that some participants behave as predicted by rational choice theory in that they never punish. The exact reasons for this behavior are not explicitly measurable in our study, but by looking at the answers to the open question in Section 7.4.5, we can specify

what considerations recipients made when deciding. Interestingly, in the first part of the experiment, the dictator game with punishment, only 31.58% of all participants do not punish, of which 5.43% award punishment points in this part. This further supports the conclusion that some participants behave as predicted by rational choice theory. However, purely self-interested rational behavior cannot explain the overall punishment behavior in this study.

Result 1 (Rationality) *Since only some recipients act in a rational and self-interested, payoff-maximizing way, the hypothesis 1 must be rejected.*

Outcome-oriented models, as proposed by Fehr and Schmidt (1999) or Bolton and Ockenfels (2000), assume that recipients punish when payoffs are distributed unequally due to people's inequality aversion. Transferring this to our study implies that recipients do not punish when an equal outcome is implemented because all group members receive the same payoff. However, the punishment behavior in this study indicates that almost 10% of the recipients assign punishment points when the equal outcome is implemented. Additionally, one could argue that in these cases, recipients only punish the decision-maker who chooses the unequal outcome, which is, however, not confirmed. Even in cases where an equal outcome is implemented for the whole group, some recipients also punish when decision-makers choose an equal allocation (6.43%) and do not distribute these punishment points equally across decision-makers, so payoff differences still exist. Consequently, Hypothesis 2 is rejected.

Result 2 (Outcome) *Purely outcome-based models alone cannot explain the punishment behavior in this study, since an equal outcome is punished almost 10% of the time, and punishment points are not evenly distributed among decision-makers.*

A different explanation for this punishment behavior is proposed by Mollerstrom et al. (2015), which specifies a different view of fairness that is agency-dependent and conditional on aspects of agents' choices, regardless of whether they matter for the outcome. The so-called 'norm choice compensation' is closely related to luck egalitarianism, but extended to situations where the outcome is not caused or influenced by the decision. It allows punishing the decision-maker for an unequal choice, even if an equal outcome is realized.

Turning to the individual level, i.e., how individual decision-makers are punished, we see that an unequal choice is punished on average with 1.89 punishment points, while an equal choice is punished with 0.14 points (two-tailed t-test, $p < 0.001$). In total, we have 525 unequal choices (21.6%) in our dataset, of which half (50.3%) are punished, while equal choices are punished in 6.4% of all cases. Impactful choices, i.e. unequal choices that can still

influence the outcome, are punished in 50.96% of all cases (mean of 1.90 points). Thus, it appears that the probability of punishment is higher for unequal choices and for impactful choices, although the difference between both motives is minimal.[57] Consequently, we can specify our next two results and confirm Hypothesis 3 and 4.

Result 3 (Choice) *Choosing the unequal allocation is punished significantly more than choosing the equal allocation.*

Result 4 (Unkind) *Revealing an unkind intention, expressed by an unequal choice that can still affect the outcome, leads to more punishment.*

To further investigate the explanatory power of unequal choices and the combined effect with other punishment motives, we refer in the regression analysis in the following sections.

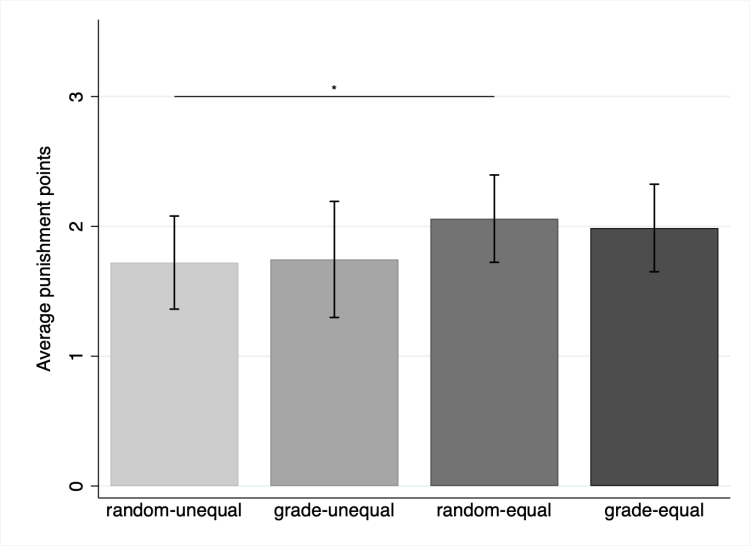
Comparing Punishment Behavior between Treatments

In terms of treatment differences, Figure 7.2 depicts the mean punishment points for an unequal choice (regardless of the implemented outcome) by treatment. As expected, participants assign slightly more punishment points to an unequal choice in the random-equal treatment than in the random-unequal treatment (one-tailed t-test, $p < 0.1$).

This is a first indication that actively choosing unequal (because the default is equal) may be more punished than passively confirming the unequal choice. Aggregating the equal and unequal default treatments (equal vs. unequal) confirms significant differences, as significantly more punishment points are assigned when the default is equal than unequal (2.02 vs. 1.73, one-tailed t-test, $p < 0.1$). In contrast, the aggregated data of the group-building mechanism (random vs. grade) do not reveal significant differences (1.89 vs. 1.89, one-tailed t-test, $p = 0.503$). Thus, the default allocation seems to have a stronger effect on the punishment decision than the mechanism.

In Section 2.1.3, we identified two mutually exclusive punishment motives: 'pivotality' and 'initiation'. While the former motive has already been elaborated theoretically (Chockler and Halpern 2004; Engl 2018) and experimentally (Bartling et al. 2015; Anselm et al. 2022), the latter motive, in particular, has been defined theoretically (Spellman 1997). Thus, it is reasonable to assume that pivotality may be more pronounced in our study. However, there are also studies focusing on proposal power (e.g. Duch et al. 2014; Duch and Stevenson

[57] This minimal difference between both motives further support to include only one motive (choice) in the following regression analyses.



Notes: Mean of assigned punishment points (from 0 to 7). The p-values are from two-tailed t-tests of treatment differences, assuming equal variances. Bars represent the 95% confidence interval. $^{*}p \leq 0.1$, $^{**}p \leq 0.05$,$^{***}p \leq 0.01$.

Figure 7.2.: Mean Punishment for an Unequal Choice by Treatment

2014), where a randomly determined decision-maker proposes an equal or unequal allocation, which has been identified as a focal point for the punishment decision. Thus, we should not underestimate the potential influence of being the first decision-maker to choose the unequal allocation. In the remainder of this section, we address this conundrum by examining the two punishment motives of 'pivotality' and 'initiation' in more detail.

Table 7.9 gives an overview of the assigned punishment points for the two different motives. The means are reported separately by treatment, default, and mechanism to examine whether the pivotality or initiation effect is present in our sample. Two main points emerge from the data in Table 7.9. The first row confirms a significant difference in the punishment points assigned between the two motives, with the initiator receiving more punishment points. Specifically, the initiator receives on average 2.30 punishment points while the pivotal decision-maker receives 1.82 points, which is significant at the 5%-level. However, this effect is not stable across treatments. For example, comparing the assigned punishment points in the two default treatments (rows two and three), we see that the initiator is assigned more punishment points (in absolute terms) when the default is equal than when the default is unequal, although this is not statistically significant. On the other hand, the initiator is punished significantly more (5%-level) when the unequal allocation is preselected (2.40 vs. 1.53).

Table 7.9.: Punishment Points for the Initiator and Pivotal Decision-maker

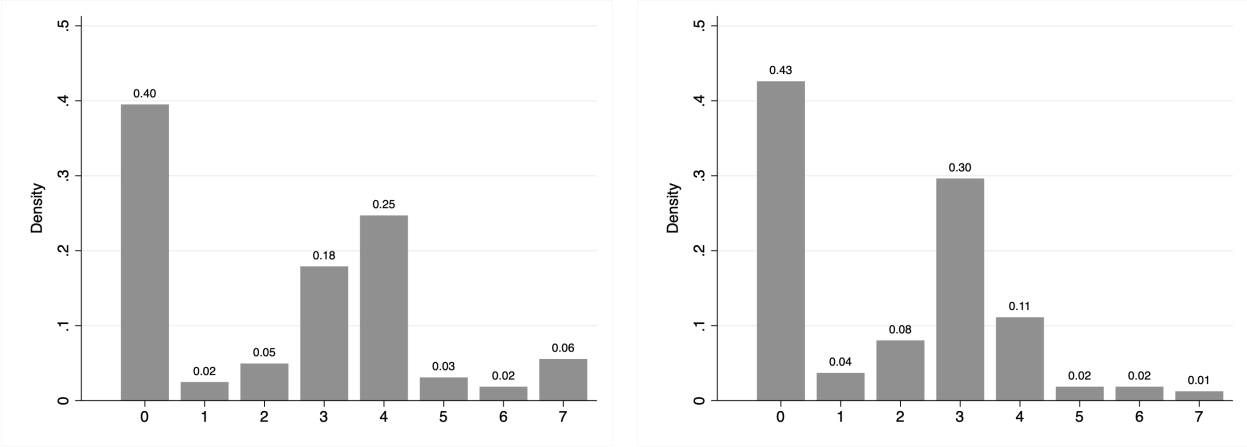
			Mean punishment points		t-test
		N	Initiator	Pivotal	Initiator vs. Pivotal
Overall		162	2.30	1.82	$p = 0.030$
Default	Unequal	57	2.40	1.53	$p = 0.021$
	Equal	105	2.25	1.98	$p = 0.332$
Mechanism	Random	75	2.37	1.96	$p = 0.169$
	Grade	87	2.24	1.70	$p = 0.095$
Treatment	Random-unequal	30	2.57	1.90	$p = 0.170$
	Grade-unequal	27	2.22	1.11	$p = 0.061$
	Random-equal	45	2.24	2.00	$p = 0.528$
	Grade-unequal	60	2.25	1.97	$p = 0.463$

Table notes: N = number of observations. Only unequal outcomes are considered. Initiator equals 1 if it is the first decision-maker choosing the unequal allocation (category (6)). Pivotal equals 1 if it is the second decision-maker choosing the unequal allocation (category (7)). P-values are from two-tailed t-tests of treatment differences, assuming equal variances. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Similarly, a significant difference in assigned punishment points is observed in the grade treatments ($p < 0.1$), while initiation and pivotality are not punished significantly differently in the random treatments (although a clear difference exists in absolute terms). Consequently, it is not surprising that a distinction by treatment detects a significant difference in assigned punishment points only in the grade-unequal treatment.

To further examine the occurrence of these two punishment motives, Figure 7.3 illustrates the distribution of punishment points for the initiator and the pivotal decision-maker. Three patterns emerge: first, participants assign almost equal numbers of zero punishment points to the pivotal decision-maker and the initiator (43% vs. 40%). Second, people are six times more likely to assign seven punishment points to the initiator than to the pivotal decision-maker. Third, when punishment points are assigned, the conditional median for the initiator is four points, while it is three points for the pivotal decision-maker. Performing a χ^2-test results in a high χ^2-value with ($\chi^2 = 19.77$, $p = 0.006$), and the p-value indicates that the distributions between the two motives are significantly different.

Interestingly, comparing the punishment behavior for the initiator and pivotal decision-maker of men and women separately reveals that these differences are more nuanced than



(a) Initiation (b) Pivotality

Notes: Mean of assigned punishment points (from 0 to 7). Initiation and pivotality for the unequal outcome as defined in Section 7.4.1.

Figure 7.3.: Initiation versus Pivotality for an Unequal Outcome

(perhaps) thought. Specifically, on average, men assign significantly more punishment points to the initiator than to the pivotal decision-maker (2.61 vs. 1.89, two-tailed t-test, $p < 0.05$). Women also punish more in absolute terms, but this is not statistically significant (2.10 vs. 1.78, two-tailed t-test, $p = 0.268$). Performing a χ^2-test leads to somewhat opposite results, as the punishment distribution for men is not significantly different between the two motives ($\chi^2 = 9.166, p = 0.241$), while it is significantly different for women ($\chi^2 = 18.014, p = 0.012$).

Overall, we conclude that participants punish the initiator significantly more than they punish the pivotal decision-maker, which is clarified in our next result. Consequently, Hypothesis 6 must be rejected, while Hypothesis 5 is accepted.

Result 5 (Initiation) *Being an initiator is punished more than being a pivotal decision-maker.*

However, these differences may be driven by treatment variation, i.e., the legitimacy of the group-building mechanism and the default, which means that the influence of both punishment motives may not be as straightforward as expected. More details and further analysis of the causal effect of different punishment motives and treatment variations follow in the next two sections, where we present our econometric comparison of different punishment motives. In particular, in Section 7.4.4 we further address this by elaborating on the joint influence of default (mechanism) and initiation (pivotality) on the assigned punishment points.

Summary 7.4: Punishment Motives

Before turning to that, we can summarize our findings concerning the punishment motives, providing an overview of hypotheses that have already been confirmed and rejected.

We rejected two hypotheses. First, Hypothesis 1 as participants do not behave entirely rationally, and second, Hypothesis 6, as the pivotal decision-maker is not punished the most. Additionally, three other hypothesis were confirmed. An unequal choice (Hypothesis 3) and an unkind intention (Hypothesis 4) lead to an increase in punishment points assigned. The initiator, thus the first decision-maker who chooses the unequal allocation, is punished the most, thereby confirming Hypothesis 5.

Recipients Expected Choices

Turning to the recipients' choices, or rather how they would decide as decision-makers, we want to examine whether they punish differently if they expect a similar decision in the corresponding decision sequence. Therefore, we first generate three dummy variables, each of which equals 1 if the expectation is confirmed. More precisely, the first belief variable equals 1 if the first recipient (B1) chooses the same allocation as the first decision-maker (A1). The second belief dummy equals 1 if the second recipient (B2) is in the same decision sequence as the second decision-maker (A2) and chooses the same allocation. Thus, belief one is true (since A1 and B1 chose the same allocation) and A2 and B2 also chose the same allocation. The third dummy variable is calculated in the same way. Belief one and two are equal to 1 to ensure that recipient three (B3) is in the same decision sequence as decision-maker three (A3), and the choice of B3 is compared to the choice of A3. After creating these dummy variables, we can compare the punishment behavior of recipients in each position (B1 through B3) with a confirmed belief, meaning that the recipient would have decided the same way as the corresponding decision-maker, and an unconfirmed belief. According to Hypothesis 12, we expect more punishment to be imposed when the recipients would have decided differently. meaning the belief is false.

We calculate t-tests for each position, shown in Table 7.10, to compare punishment points between confirmed and unconfirmed beliefs. Overall, recipients' punishment behavior is strongly influenced by their expectations of how a decision-maker might decide in the same decision sequence. Specifically, the first recipient (B1) awards an average of 0.90 punishment

Table 7.10.: Punishment Behavior Dependent on Belief of Recipients

		B1		B2		B3	
		Mean	t-test	Mean	t-test	Mean	t-test
Belief	False	0.895	$p < 0.001$	1.021	$p < 0.000$	0.731	$p < 0.000$
	True	0.315		0.331		0.140	

Table notes: Mean of assigned punishment points (from 0 to 7). Number of observations: 270. Belief is true if the recipient would have decided the same way as the corresponding decision-maker. The p-values are from two-tailed t-tests, assuming equal variances. $^{*}p \leq 0.1$, $^{**}p \leq 0.05$,$^{***}p \leq 0.01$.

points when the decision-maker actually chooses a different allocation than the recipient would expect. On the other hand, if the decision-maker and the recipient choose the same allocation, 0.32 punishment points are assigned (two-tailed t-test, $p < 0.01$). For recipients two (B2) and three (B3), the same punishment behavior is observed, as significantly fewer punishment points are assigned when the decision-maker chooses the same allocation as the recipients would expect.

Consequently, the beliefs of the recipients, or rather how they expect the decision-makers to decide, significantly affect their punishment decision. In Section 7.4.4, we generate a dummy variable from the three previously presented variables and include it as an additional control variable in our analyses.

7.4.3. Determinants of the Decision to (not) Punish

In this section, we begin by estimating probit regressions to examine the probability that a recipient punishes the decision-maker where the exact amount of punishment points is not relevant. The punishment variable is coded as a dummy variable equal to one if at least one punishment point is assigned. First, the mechanism (Model 1) and the default (Model 2) are regressed separately, and then their interaction, i.e., the treatment distinction, is regressed (Model 3). In Model 4, a dummy variable *choice default* is added, which equals 1 if the default option is chosen. Then, several control variables (Model 5), more precisely the characteristics of the decision-maker and the rating results, are included. In order to have a clear distinction between the two motives 'initiation' and 'pivotality', we decide to include only the initiator of the unequal outcome ('outcome X initiator') in our regression analyses, since participants are classified as pivotal for the unequal outcome. Including both

variables, 'initiator' and 'outcome X initiator', leads to contradictory results because both variables are highly correlated (0.65).[58] This allows for the disentangling of potential effects and for a better understanding of whether the punishment decision is driven by the default, the mechanism, a combination of both, or other components. Finally, Model 7 extends the sixth model by including the four subscales of the BSJO scale, which represent stated preferences for the four main distributive justice criteria. Marginal effects are shown because their interpretation is more meaningful.[59]

The effect of mechanism and default is addressed in models one through three. Without additional control variables, the group-building mechanism does not significantly affect the decision to punish, while an equal default increases the probability of being punished by about seven percentage points ($p < 0.05$). Thus, the preselected equal allocation, where the unequal allocation has to be actively implemented, leads to a higher probability of punishment. Differentiating by treatment (Model 3) indicates that the influence of default almost disappears. There is only a weakly significant difference between random-unequal (the baseline) and grade-equal, where the probability of punishment is 4.3 percentage points higher ($p < 0.1$). However, including the default choice dummy (Model 4) attenuates this weak effect, with no significant treatment differences.

Extending the model with different control variables (Model 5) provides further insights. The most striking finding is that we still do not have significant treatment effect and that a choice of the default has no effect on the probability of being punished. Not surprisingly, however, the probability to punish is significantly lower for women (-9.2 percentage points, $p < 0.05$). In addition, there is a highly significant negative effect of belief, i.e., how a recipient expects a decision-maker to decide in the corresponding decision sequence. More specifically, if the recipient expects the same decision as the decision-maker actually made, the probability of punishment decreases by about eight percentage points ($p < 0.01$). Random's legitimacy rating also influences the probability of punishment: for one point increase on the rating scale, the probability decreases by 1.7 percentage points ($p < 0.1$).

Model 6 includes the different punishment motives and confirms the previous finding of no significant treatment differences. However, an unequal choice is highly influential in the punishment decision, as it increases the probability of punishment by about 30 percentage points ($p < 0.01$). Interestingly, being the initiator and being the pivotal decision-maker both

[58] Nevertheless, we also estimate the regressions with 'initiation' instead of 'outcome X initiator', which leads to a higher AIC value, indicating that a model with 'outcome X initiator' fits better, although both lead to the same significant results.

[59] In Appendix A.4 a correlation matrix with all variables is displayed. As expected, punishment motives are (strongly) correlated, while the correlation between the other variables is acceptable.

Table 7.11.: Probit Regression - Marginal Effects

Punishment - Dummy	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mechanism	-0.024						
	(0.032)						
Default		0.068**					
		(0.027)					
Grade-unequal			-0.029	-0.029	-0.014	-0.023	-0.013
			(0.021)	(0.023)	(0.018)	(0.028)	(0.026)
Random-equal			0.065	0.086	0.115	0.070	0.071
			(0.050)	(0.080)	(0.085)	(0.046)	(0.044)
Grade-equal			0.043*	0.062	0.059	0.010	0.014
			(0.025)	(0.054)	(0.051)	(0.024)	(0.028)
Choice default				-0.047	-0.045	0.029	0.029
				(0.107)	(0.105)	(0.024)	(0.024)
Female					-0.092**	-0.075**	-0.084*
					(0.037)	(0.037)	(0.044)
Age					0.002	0.003	0.004
					(0.005)	(0.003)	(0.003)
Risk affine					-0.006	0.024	0.020
					(0.053)	(0.040)	(0.033)
Belief					-0.079***	-0.014	-0.013
					(0.019)	(0.020)	(0.020)
Economist					-0.039	-0.043	-0.038
					(0.039)	(0.036)	(0.030)
Deserve role A					-0.006	-0.004	-0.009
					(0.008)	(0.007)	(0.007)
Deserve role B					-0.010	-0.005	-0.005
					(0.009)	(0.008)	(0.007)
Rating random					-0.017*	-0.014	-0.014
					(0.010)	(0.008)	(0.009)
Rating grade					0.005	0.012*	0.011
					(0.008)	(0.006)	(0.007)
Outcome unequal						0.024	0.025
						(0.024)	(0.025)
Choice unequal						0.217***	0.218***
						(0.024)	(0.025)
Outcome X Initiator						0.066***	0.067***
						(0.021)	(0.021)
Pivotal						0.050***	0.051***
						(0.018)	(0.018)
BSJO equity							0.018
							(0.017)
BSJO equality							-0.011
							(0.023)
BSJO need							0.047
							(0.031)
BSJO entitlement							0.008
							(0.018)
Wald-χ^2	0.59	7.37	45.54	46.31	/	/	/
$p(\chi^2)$	0.441	0.007	0.000	0.000	/	/	/
AIC	1637.25	1634.18	1637.613	1626.85	1605.99	852.65	855.09
N	2430	2430	2430	2430	2430	2430	2430

Table notes: Dependent variable: Punishment dummy which equals 1 if at least one punishment is assigned. Average marginal effect of a random effects probit regression with standard errors in parentheses and clustered on session. Mechanism equals 1 if grade is the group-building mechanism. Default equals 1 if the preselected option is equal. Random-unequal is the baseline category. Choice default equals 1 if the preselected option is chosen. Belief equals 1 if the recipient would have decided the same way as the corresponding decision-maker. The variables 'Female', 'Economics', 'Risk affine' are dummy variables for the respective socio-demographic variable. Deserve Role A and B range from 1 (completely not deserving) to 10 (completely deserving). Rating random and grade range from 1 (completely not legitimate) to 7 (completely legitimate). Punishment motives are dummy variables for the respective category as defined Section 7.4.1. $***p < 0.01, **p < 0.05, *p < 0.1$.

increase the probability of punishment, but the increase is higher for being the initiator. More specifically, being the initiator leads to an increase of 6.6 percentage points ($p < 0.01$), while being pivotal increases the probability of punishment by 5.0 percentage points ($p < 0.01$). Again, the female dummy has a negative sign and is statistically significant in the last two models (Model 6 and 7), indicating that the probability of punishment for women is 7.5 percentage points lower in Model 6 ($p < 0.05$) and 8.4 percentage points lower in Model 7 ($p < 0.1$). Furthermore, contrary to the results of Model 5, for a one point increase in legitimacy rating of the final grade, the probability of punishment is 1.2 percentage points higher in Model 6 ($p < 0.1$). In contrast, the legitimacy rating of *random* is no longer significant. Interestingly, there is no effect of deservingness, so the evaluation does not influence the decision to punish. Furthermore, belief, or how a recipient would expect the decision-maker to decide, no longer influences the decision to punish.

Finally, in Model 7, we include the BSJO scale, or rather the four subscales, each of which represents one of the main principles of justice. We do not gain further insight into when participants punish because the significant and marginal effects between the last two models are stable. Therefore, we will not discuss the BSJO scale in detail here, but will return to it in the next section. According to the AIC values, the sixth model is the best, as it has the lowest AIC value.

Summary 7.5: Results of Probit Regressions

So far, the probit regressions suggest that the group-building mechanism does not seem to influence the decision to punish, while the default does. However, when both variations are combined, there are only weakly significant treatment differences, which disappear entirely when additional control variables and punishment motive are added.

In addition, the probability to punish is lower for women than for men, which is stable across the estimated models. An unequal choice increases the likelihood of punishment the most, although being the initiator and being the pivotal decision-maker also increase the likelihood of punishment.

In the following section, we further investigate the effect of these factors on the (exact) amount of punishment points assigned. Thus, having identified the factors that influence the decision to punish at the extensive margin in this section, we now turn to the effects at the intensive margin by focusing on the amount of punishment points assigned.

7.4.4. Assignment of Punishment Points

Next, we turn to the tobit regression results, following the same pattern as for the probit regression in the previous section. We begin by regressing the treatment variations (default, mechanism, and their interaction) on the number of punishment points assigned. We then include control variables and punishment motives, as well as the BSJO scale, to examine their influence on the number of punishment points assigned. Because the linear dependent variable (punishment points) is heavily skewed to the left, a tobit model is a better fit than an OLS regression.[60]

Models 1 and 2 in Table 7.12 estimate the effect of the mechanism and the default, respectively, on the punishment points assigned. Similar to the results of the probit regression, the mechanism does not significantly determine the assigned punishment points, while the default influences the punishment behavior. When an equal allocation is preselected, participants assign on average 0.2 more punishment points than with an unequal default ($p < 0.1$). This is not surprising, since an equal default requires an active choice of the unequal (unfair) allocation, which is expected to lead to more punishment (see Hypothesis 7). Turning to the interaction of default and mechanism in Model 3, i.e., the treatment distinction, there are no significant differences between treatments, which is confirmed by mutual Wald F-tests for significant differences between the four treatments (not shown in Table 7.12).

Model 4 further explores the impact of default by including a dummy variable that equals one if the default is chosen, which reveals some interesting patterns. First, choosing the default results in significantly fewer punishment points (0.25 points), which is highly significant ($p < 0.01$). In addition, the inclusion of the dummy variable leads to significant treatment differences. In random-equal and grade-equal, participants assign on average 0.32 more punishment points, 0.34 respectively, than in random-unequal ($p < 0.05$). However, since this effect is only apparent between unequal and equal treatments, the previously discussed default effect is likely driving this punishment decision. Restricting the observations in Model 4 to only unequal outcomes (not reported in Table 7.12), this effect is even more pronounced. When the unequal outcome is implemented, choosing the default is punished significantly less (0.65 points, $p < 0.001$) than not choosing the default. We will come back to the default effect in Section 7.4.4.

The next two models include various control variables (Model 5) and punishment motives (Model 6). In Model 5, we first add socio-demographic characteristics and find that women punish significantly less (-0.24 points, $p < 0.05$) than men do. This is consistent with the

[60] Nevertheless, we estimate an OLS regression, which is presented in Table A.5 in Appendix A.4.

Table 7.12.: Tobit Regression

Punishment points	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mechanism	0.008						
	(0.106)						
Default		0.201*					
		(0.104)					
Grade-unequal			-0.018	-0.025	0.060	0.066	0.108
			(0.146)	(0.145)	(0.147)	(0.146)	(0.142)
Random-equal			0.179	0.320**	0.406***	0.215	0.217
			(0.146)	(0.150)	(0.142)	(0.140)	(0.136)
Grade-equal			0.206	0.337**	0.353**	0.109	0.137
			(0.146)	(0.149)	(0.148)	(0.146)	(0.142)
Choice default				-0.246***	-0.248***	-0.050	-0.050
				(0.062)	(0.062)	(0.051)	(0.051)
Female					-0.239**	-0.212*	-0.258**
					(0.110)	(0.109)	(0.110)
Age					0.012	0.016*	0.020**
					(0.009)	(0.009)	(0.009)
Risk affine					-0.092	-0.008	-0.004
					(0.124)	(0.123)	(0.119)
Belief					-0.246***	-0.033	-0.033
					(0.065)	(0.053)	(0.053)
Economist					-0.186	-0.224*	-0.213*
					(0.126)	(0.125)	(0.123)
Deserve role A					-0.001	0.004	-0.013
					(0.025)	(0.025)	(0.025)
Deserve role B					-0.032	-0.018	-0.015
					(0.023)	(0.023)	(0.023)
Rating random					-0.054*	-0.044	-0.047
					(0.030)	(0.029)	(0.029)
Rating grade					-0.009	0.014	0.011
					(0.031)	(0.031)	(0.032)
Outcome unequal						-0.081	-0.077
						(0.083)	(0.083)
Choice unequal						1.533***	1.532***
						(0.073)	(0.073)
Outcome X Initiator						0.750***	0.751***
						(0.127)	(0.127)
Pivotal						0.266**	0.267**
						(0.127)	(0.127)
BSJO equity							0.052
							(0.062)
BSJO equality							-0.083
							(0.056)
BSJO need							0.186**
							(0.083)
BSJO entitlement							0.008
							(0.071)
Constant	0.513***	0.419***	0.427***	0.477***	0.852***	0.083	-0.653
	(0.074)	(0.073)	(0.101)	(0.101)	(0.316)	(0.314)	(0.584)
Wald-χ^2	0.01	3.71	3,76	19.40	54.00	1375.71	1387.03
$p(\chi^2)$	0.939	0.054	0.289	0.001	0.000	0.000	0.000
AIC	8032.58	8028.93	8032.88	8019.32	8005.09	6977.57	6976.81
N	2430	2430	2430	2430	2430	2430	2430

Table notes: Dependent variable: Punishment points (from 0 to 7). Random effects tobit regression with standard errors in parentheses. Mechanism equals 1 if grade is the group-building mechanism. Default equals 1 if the preselected option is equal. Random-unequal is the baseline category. Choice default equals 1 if the preselected option is chosen. Belief equals 1 if the recipient would have decided the same way as the corresponding decision-maker. The variables 'Female', 'Economics', and 'Risk affine' are dummy variables for the respective socio-demographic variable. Deserve Role A and B range from 1 (completely not deserving) to 10 (completely deserving). Rating random and grade range from 1 (completely not legitimate) to 7 (completely legitimate). Punishment motives are dummy variables for the respective category as defined in Section 7.4.1. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$..

estimated probability of punishment, which is also significantly lower for women. In addition, the significant treatment differences and the effect of a default choice are still visible. As in the corresponding probit regression, belief is highly statistically significant, implying that recipients punish less (-0.25 points, $p < 0.01$) when they expect the same allocation as decision-makers actually choose.

However, the inclusion of punishment motives changes this picture (Model 6), as choosing the default no longer significantly increases the assigned punishment points. However, the coefficient is still negative. In addition, significant treatment differences disappear, suggesting that punishment motives may have a stronger effect on assigned punishment points than treatment differences have. In other words, in predicting punishment behavior, the influence of treatment differences is mitigated by punishment motives, which will be discussed further in the next section.

In this respect, all punishment motives, except for unequal outcome, lead to a significant increase in the assigned punishment points. For an unequal choice, the assigned punishment points are on average 1.53 points higher ($p < 0.01$) than for an equal choice. In addition, the initiator receives 0.75 more punishment points ($p < 0.01$), while the pivotal decision-maker receives 0.27 more punishment points ($p < 0.05$). A Wald F-test confirms that both motives, 'initiation' and 'pivotality', are significantly different from each other ($\chi^2 = 20.58, p < 0.001$).

Finally, in Model 7, we include the BSJO-scale. As a reminder, the BSJO-scale measures individuals' perceptions of the four main principles of distributive justice: equality, equity, need, and entitlement.[61] It consists of eight items and has already been used in three general population surveys in Germany.[62] As it relates to rules and norms that people might consider relevant in allocation decisions, it fits quite well into the overall context of this study. By including participants' responses, we have information about their stated preferences, which may not match their revealed punishment preferences. In addition, as it is already used in population surveys, we can compare the attitudes in our study with those of the general

[61] For more information on the principles, see Section 3.3.3. The exact wording of the four dimensions and eight scale items is listed in Table A.8 in Appendix A.4. No translation was necessary as the scale was developed in German and has been used in German-speaking contexts. The exact wording of the items is taken from LINOS-1, which is recommended by the authors (Hülle et al. 2018, p. 672). Prior to the analysis, the values of each item must be converted so that they are comparable across datasets, with higher values reflecting higher agreement. The German Data Forum (Rat für Sozial- und Wirtschaftsdaten) recommends quality standards for survey experiments regarding the objectivity, reliability, and validity of scales, which were used by Hülle et al. (2018). Therefore, in order to examine and demonstrate the quality of the BSJO scale in this sample, the same tests are conducted with this dataset (see Appendix A.4).

[62] More specifically, it has been implemented in "The Legitimation of Inequality Over the Life Span" (LINOS-1), the Innovation Sample of the German Socio-Economic Panel (SOEP-IS 2012), and the German Social Survey (ALLBUS 2014).

population.

A comparison among the participants in this study (students at the University of Hamburg) and the general population is provided in Table A.9 in the Appendix A.4. It indicates that the attitudes of the two samples are quite similar.[63] Next, we use the four subscales, each representing a justice principle, and include them in our tobit regression (Model 7).

As it can be seen in Table 7.12, there are no treatment effects and the coefficients and significance levels are quite similar between Models 6 and 7. When the four subscales are combined, there is a significant positive effect of need, which is due to multicollinearity. Interestingly, entitlement as a stated justice attitude is not significantly correlated with subjects' punishment behavior, nor is the equality criterion, according to which everyone should receive the same payoff, although most participants state equality as a criterion influencing their punishment decision. We will come back to this point in Section 7.4.5, when presenting the results of the open question. Since both criteria may be more pronounced for unequal outcomes, we estimate the same tobit regression considering only observations where an unequal outcome is implemented (see Table A.7 in Appendix A.4, column 7). However, even with unequal outcomes, neither entitlement nor equality oriented participants punish more. Interestingly, participants assign significantly more punishment points (10%-level) when they say they agree more with the equity principle, according to which everyone should receive the amount proportional to their input. This may be an indicator that participants perceive the input of all participants to be similar, so that everyone should receive the same amount, resulting in more assigned punishment when no equal distribution is implemented. Thus, the situational context or rather the unfairness of the implemented results seems to be a factor.

Since the dummy variable 'risk affine' is far from being statistically significant in Models 5 through 7, we conclude that risk attitude does not determine the amount of punishment points assigned.[64] Combined with the results from the probit regressions, where risk attitude also does not affect the probability of punishment, we can formulate our next result and reject Hypothesis 11.

Result 6 (Risk Attitude) *In our setting, the individual's risk preference does not affect the probability of punishment or the amount of punishment points assigned.*

Finally, we compare the AIC values of the seven models to see which one fits best. The

[63] The results of LINOS-1 are used because this survey is the benchmark and determines the exact wording of the items.

[64] Using a dummy for being risk averse, so that the comparison group consists of risk affine and risk neutral participants, confirms these results.

AIC values of Models 6 and 7 are the lowest and quite similar, leading to the conclusion that a model with punishment motives is more appropriate than models without punishment motives.

Taking the results of the probit and tobit regressions together, we can specify our next result on gender differences in punishment.

Result 7 (Gender) *Punishment behavior is influenced by gender in that being a female negatively affects the probability of punishment and the amount of punishment points assigned. As a result, we confirm our Hypothesis 10.*

Before turning to a more detailed examination of a (potential) default and mechanism effect as well as further robustness checks, we summarize our results as follows:

Summary 7.6: Results of Tobit Regressions

The default alone has a significant effect on the amount of punishment points assigned, while treatment differentiation only reveals significant punishment differences as long as punishing motives are not considered. A choice of the default has a significant negative effect in Models 4 and 5, which disappears when we control for the different punishment motives and the BSJO-scale. Thus, the punishment motives determine the assigned punishment points and not the default choice or the treatment differences. In addition, females and economists assign significantly fewer punishment points.

Testing for a Default Effect

Recipients may perceive the default as the recommended option, and thus punish less harshly when the default is chosen. In all cases, 46.3% of the choices are default choices, while the choices differ significantly by default (see Section 7.2). Table 7.12 shows that the negative effect of choosing the default option on the allocation of punishment points disappears when the different punishment motives are included (Model 6 and 7). In this section, we further investigate whether the punishment behavior of the initiator and the pivotal decision-maker depends on the default and whether they are punished less (more) for an unequal choice when it is the default option. More specifically, we want to examine whether there is a relationship between being the initiator (or pivotal) and actively or passively implementing the unequal option.

Table 7.13.: Testing for a Default Effect

Punishment points	(8)	(9)	(10)	(11)
Mechanism	-0.017	0.014	-0.013	0.013
	(0.101)	(0.104)	(0.101)	(0.104)
Default - Initiator				
Unequal - No	-0.144	-0.118		
	(0.102)	(0.097)		
Unequal - Yes	1.930***	0.855***		
	(0.180)	(0.186)		
Equal - Yes	1.726***	0.658***		
	(0.117)	(0.139)		
Default - Pivotal				
Unequal - No			-0.127	-0.075
			(0.102)	(0.097
Unequal - Yes			0.983***	-0.028
			(0.185)	(0.186)
Equal - Yes			1.419***	0.420***
			(0.122)	(0.139)
Constant	0.475***	-0.520	0.501***	-0.539
	(0.088)	(0.588)	(0.089)	(0.587)
Punishment motives	No	Yes	No	Yes
Controls	No	Yes	No	Yes
Wald-χ^2	398.01	1389.91	185.52	1392.03
$p(\chi^2)$	0.000	0.000	0.000	0.000
AIC	7670.15	6973.16	7859.71	6971.76
N	2430	2430	2430	2430

Table notes: Dependent variable: Punishment points (from 0 to 7). Random effects tobit regression with standard errors in parentheses. Mechanism equals 1 if grade is the group-building mechanism. Equal default and not being the initiator (pivotal, respectively) as baseline category. Control variables and punishment motives as in Model 7 with punishment motives as defined in Section 7.4.1. $***p < 0.01, **p < 0.05, *p < 0.1$.

Therefore, we estimate four additional models in which we include the interaction of default and initiator (pivotality, respectively). Since the punishment motive 'initiation' already implies an unequal choice, we interact this with the preselected option to see if an unequal choice that could have been passively implemented is less punished. Additionally, for each interaction, we estimate a model without the control and punishment variables and a model with them included. The baseline category is an unequal default and not being the initiator

or the pivotal decision-maker.

The first two columns in Table 7.13 show the interaction between the default and being the initiator, uncovering some interesting patterns. Being the initiator, i.e., being the first decision-maker to make an unequal choice, leads to significantly more punishment points, regardless of the default. However, compared to an equal default and not being the initiator, an initiator is punished 1.93 points more with an unequal default ($p < 0.01$) and 1.726 points more with an equal default ($p < 0.01$). Adding control variables and punishment motives (Model 9) reduces the increase in assigned punishment points by about one point, although the effect is still present and highly significant. Interestingly and unexpectedly, the increase in assigned punishment points is higher when the unequal allocation is preselected, so the unequal allocation can be passively implemented.

The interaction between pivotality and the default option is somewhat different. Without control variables (Model 10), pivotality is punished significantly more than non-pivotality, regardless of the default. However, pivotality combined with an equal default, so that the unequal allocation has to be chosen actively, is punished more than implementing the unequal allocation passively (Coef. 1.419 vs. 0.983, $p < 0.01$). Adding the control variables and the other punishment motives reveals that the significant effect of being pivotal with an unequal default, compared to not being pivotal with an equal default, disappears. The interaction of pivotality with an equal default is still highly significant (Coef. 0.42, $p < 0.01$).

To further explore the relationship between initiation, pivotality, and default, we look at the marginal effects corresponding to Models 9 and 11, reported in Table 7.14.

Table 7.14.: Testing for a Default Effect - Marginal Effects

		Initiation		Pivotality	
		No	Yes	No	Yes
Default	Unequal	0.408***	1.381***	0.461***	0.508***
		(0.067)	(0.169)	(0.067)	(0.169)
	Equal	0.526***	1.180***	0.536***	0.956***
		(0.069)	(0.146)	(0.068)	(0.145)

Table notes: Average marginal effect of tobit regressions (Model 9 and 11) with standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$.

The marginal effects provide a more nuanced picture of punishment behavior. Regarding the punishment of the initiator, the marginal effects are as expected. The initiator receives more punishment points with an unequal default than with an equal default. Thus, implementing the unequal allocation actively (equal default) or passively (unequal default) makes

only a small difference in the assigned punishment points, in favor of an active implementation, which is quite unexpected. On the other hand, choosing the default can prevent punishment or lead to a lower amount of assigned punishment points if the decision-maker is pivotal. The marginal effects of pivotality are lower for an unequal default than for an equal default (0.508 versus 0.956).

Summary 7.7: Testing for a Default Effect

Consequently, choosing the default does not prevent you from being punished as a first decision-maker, since the initiator is punished significantly more, regardless of the default. However, choosing unequal and being the pivotal decision-maker with an unequal default is punished less severely (in terms of absolute punishment points assigned) than with an equal default.

Taken these results together with the tobit regressions (Table 7.12), it seems that the punishment behavior towards the initiator is independent of the default, while the effect of the motive 'pivotality' is correlated with or influenced by the default. Consequently, choosing the default is less punished when no other punishment motives are included. However, adding punishment motives almost completely mitigates the default effect, except for pivotality, where a default choice can decrease the assigned punishment points. The conclusion that punishment motives are more influential than the preselected option is appropriate, leading to a rejection of Hypothesis 7.

Result 8 (Default Choice) *Choosing the default results in less punishment if no further punishment motives are controlled. However, a default choice cannot prevent punishment if the decision-maker is the initiator, while a pivotal decision-maker can mitigate the punishment points assigned by choosing the default option.*

Testing for an Effect of Status Differences

Although no effect of the group-building mechanism has been found so far, the evaluation of both mechanisms seems to occasionally influence the punishment decision. If the legitimacy of a random procedure is rated one point higher, 0.05 fewer punishment points are assigned (Table 7.12, Model 5). Although this effect disappears when controlling for punishment motives, we want to further explore whether the mechanism influences the punishment decision. Therefore, two dummy variables are generated based on the assessment of the group-building mechanism. 'Random legitimate' equals 1 if random is considered more legitimate, while 'Grade legitimate' equals 1 if the final grade is considered more legitimate.

The vast majority (75.56%) perceive random as more legitimate, while only 9.63% rate the final grade as more legitimate than a random group-building mechanism. The rest consider both mechanisms equally legitimate.

Table 7.15 displays the results of our regression analyses, where we include the two dummy variables one at a time and estimate a model with control variables (mainly sociodemographic characteristics) and punishment motives. [65] In addition, we include the interaction of being the initiator (pivotality, respectively) with rating a random procedure as more legitimate.[66]

The inclusion of the two dummy variables (Models 12 and 13) does not lead to significant treatment differences in punishment behavior. More precisely, even controlling for the evaluation of the more or less legitimate mechanism does not induce significant differences in the punishment points assigned between treatments.

Turning to the interaction of initiation and the evaluation of random (Model 14), we again observe that recipients punish the initiator significantly more, regardless of their evaluation. However, when random is rated as more legitimate than the final grade, the initiator receives fewer punishment points than when both mechanisms are rated equally or more legitimately. Specifically, a receiver assigns 1.68 more punishment points ($p < 0.01$) to the initiator when random is not rated more legitimate than to the non-initiator. Being the initiator and rating random as more legitimate increases the assigned punishment points by 0.70 ($p < 0.01$). Turning to the interaction between pivotality and the rating of random (Model 15), a different picture emerges. Being pivotal is not punished more regardless of the rating, except in the case where random is rated as more legitimate (Coef. 0.40, $p < 0.1$). Marginal effects (see Table A.6 in the Appendix A.4) confirm these results, highlighting that the initiator is punished less intensely when the recipient perceives random as more legitimate, while the effect is reversed for a pivotal decision-maker.[67]

[65] We also check for other channels, such as the interaction between the actual mechanism and the evaluation of the more legitimate mechanism (not shown in Table 7.15), which do not lead to significant results of the group-building mechanism on the assigned punishment points.

[66] We omit the interaction between grade more legitimate and initiation (pivotality, respectively) because the results are almost contradictory.

[67] As a further control, we split the dataset by mechanism and analyze the 'random' and 'grade' treatments separately, which confirms our results.

Table 7.15.: Testing for a Mechanism Effect

Punishment Points	(12)	(13)	(14)	(15)
Grade-unequal	0.107	0.103	0.108	0.107
	(0.142)	(0.143)	(0.142)	(0.142)
Random-equal	0.183	0.182	0.180	0.183
	(0.133)	(0.134)	(0.132)	(0.133)
Grade-equal	0.104	0.106	0.108	0.104
	(0.139)	(0.139)	(0.139)	(0.139)
Random legitimate	0.121			
	(0.195)			
Grade legitimate		-0.063		
		(0.225)		
Random more legitimate - Initiator				
Yes - No			0.206	
			(0.195)	
No - Yes			1.681***	
			(0.191)	
Yes - Yes			0.699***	
			(0.233)	
Random more legitimate - Pivotal				
Yes - No				0.123
				(0.196)
No - Yes				0.302
				(0.193)
Yes - Yes				0.404*
				(0.234)
Punishment motives	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Constant	-0.771	-0.644	-0.884	-0.773
	(0.608)	(0.590)	(0.606)	(0.608)
Wald-χ^2	1386.05	1385.64	1448.52	1386.06
$p(\chi^2)$	0.000	0.000	0.000	0.000
AIC	6977.39	6977.70	6940.15	6979.38
N	2430	2430	2430	2430

Table notes: Dependent variable: Punishment points (from 0 to 7). Random effects tobit regression with standard errors in parentheses. Random not more legitimate and not being the initiator (pivotal, respectively) as baseline category. Random legitimate equals 1 if random is evaluated as more legitimate than grade. Grade legitimate equals 1 if grade is evaluated as more legitimate than random. Control variables and punishment motives as in Model 7 with punishment motives as defined in Section 7.4.1. $***p < 0.01, **p < 0.05, *p < 0.1$.

Consequently, as summarized in Result 9, it is likely that the legitimacy of the group-building mechanism does not influence punishment behavior and is not used as a focal point for assigning responsibility. Contrary to the theoretical prediction, participants do not assign responsibility based on induced (legitimate) status differences, so Hypothesis 8 must be rejected.

Result 9 (Mechanism) *The empirical legitimacy of the group-building mechanism does not determine the allocation of punishment points by recipients.*

Additionally, after presenting the probit and tobit regressions, we can conclude, that there are no significant differences between treatments, i.e., the interaction of default and mechanism. Consequently, punishment behavior is not influenced by the interaction of default and mechanism, but rather by punishment motives, leading to a rejection of Hypothesis 9.

Result 10 (Interaction) *Punishment behavior is not significantly affected by the legitimacy of the group-building mechanism combined with an equal or unequal default allocation.*

7.4.5. Further Analyses

Robustness Checks

In order to strengthen the previous results, seven additional models are estimated based on Model 7 (see Table A.7 in the Appendix A.4). In the following, we describe each of these models, explain why we chose to estimate it, and present the main results, focusing on the differences with Model 7 in Table 7.12.

a) Only round 1

Since it can be argued that there is some learning effect between rounds, participants can form expectations or use the previous allocation as a reference point. Thus, a model using only the first round (since no learning or reference is possible) is estimated.
When we look at the first round only, there is a significant difference between random-unequal and random-equal, as participants punish significantly more (0.36 points, $p < 0.1$) in random-equal, which may be driven by the default. In addition, an unequal choice is punished more (2.1 points, $p < 0.01$) and the initiator receives significantly more punishment points (0.86 points, $p < 0.01$), while being pivotal does not significantly increase the assigned punishment points and has a negative sign.

b) Without round 1

Participants may be unfamiliar with the punishment decision, so the first round can

be considered an introductory or learning round. Thus, a model is estimated without the first round.

Without the first round, the economist and female dummies, as well as age, significantly influence the punishment points assigned. Specifically, economists and females assign significantly fewer punishment points (Coef. -0.26 and -0.27, respectively, $p < 0.05$), while the older the recipient is, the more punishment points are assigned (Coef. 0.02, $p < 0.05$). There are no other differences to Model 7.

c) Without last round

The last round is excluded in this estimation to control for a potential end-round effect, where participants might behave more selfishly because there is no further interaction. In contrast to Model 7, excluding a potential end-game effect reveals significant differences between random-unequal and random-equal, as participants punish significantly more in the latter (Coef. 0.33, $p < 0.05$).

d) Without seven punishment points

A model without the maximum number of punishment points is estimated to exclude the extreme value where only one decision-maker is punished.

Even without the extreme value of seven punishment points, the previously presented results are confirmed, supporting the robustness of our findings.

e) Only male and f) only female

Previously, some gender differences were found, so a separate regression for males and females is reasonable. This should strengthen the previous findings that punishment decisions are not driven by gender differences.

Regarding the two punishment motives 'choice' and 'initiation', no gender differences are found, because they are highly significant in both models. However, when we look only at the *male participants*, significant treatment differences emerge. Specifically, males assign more punishment points in grade-unequal and random-equal than in the baseline treatment random-unequal (Coef. 0.46, $p < 0.1$ and 0.59, $p < 0.05$ respectively). In addition, choosing the default is significantly less punished (Coef. -0.22, $p < 0.05$), while being the pivotal decision-maker does not significantly determine the assigned punishment points. Although the effect is weak, men's expectation of how a decision-maker might decide negatively affects the punishment points assigned (Coef. -0.17, $p < 0.1$).

In contrast, there are no significant treatment differences for women (as in Model 7),

which is not surprising since we have far more women than men in our setting. However, being pivotal affects the punishment points assigned by 0.31 ($p < 0.1$).

g) Only unequal outcome

Since the punishment for an unequal outcome is significantly higher, a robustness check focuses on the unequal outcomes to further investigate whether some effects exist only in this case.

Again, there are no treatment differences. However, being the pivotal decision-maker no longer significantly affects the punishment points assigned. In other words, when the implemented outcome is unequal, an unequal choice and being the initiator determine the assigned punishment points (Coef. 1.19, $p < 0.01$ and 1.06, $p < 0.01$, respectively).

Summary 7.8: Robustness Checks

The robustness checks lead to the conclusion that our results are robust. Specifically, the effect of the different punishment motives is confirmed in all of the further estimated models. Thus, our assumption that an unequal choice, being the initiator, and being pivotal determine the punishment decision is confirmed, although the effect of pivotality is not as robust as for the other motives. In addition, the robustness checks reveal interesting insights into the punishment behavior of men and women, as men seem to be more influenced by the treatment variations, which may be related to male competitiveness (among others, see e.g. Campbell 2002). Furthermore, pivotality does not significantly influence the punishment points assigned by men.

Since men seem to punish somehow differently than women in our setting, it would be interesting to see whether an equal sample of men and women would change our results in Model 7. Alternatively, the results may imply that there are systematic differences in punishment behavior between men and women in a setting like ours. Future research should shed light on this aspect, as the empirical evidence is not yet clear or large enough to draw conclusions about gender differences in punishment.

Individual Answers to the Open Question

In this section, we present the individual responses to the open question with the aim of clarifying whether participants really do make these considerations. More precisely, since the punishment decision reveals participants' true preferences, it is interesting to know whether they also cite these preferences as their reason for (not) punishing.

At the end of the experiment, participants are asked an open question, depending on their role. Decision-makers (role A) explain their reasoning behind their decision. Recipients (role B) explain why (and how) they assigned punishment points. The exact wording of the open question is as follows:

> *For A: In this experiment you acted as participant A and chose an allocation. What considerations did you make? /*
> *And for B: In this experiment you acted as participant B and assigned punishment points. What considerations did you make?*[68]

The top five responses are listed separately for roles A and B in Table 7.16. Fairness and equal distribution are important motives for decision-makers, while almost a quarter try to maximize their own payoff. In addition to self-interest, a quarter of decision-makers are concerned about the likelihood of being punished for their decision, while 16.13% are concerned about being punished.

Table 7.16.: Top-5 Answers to the Open Question

	Role A		Role B	
1.	40.32%	fairness	28.24%	punishment costs / is expensive
2.	25.81%	probability to be punished	23.66%	unequal choice
3.	24.19%	utility maximization	21.37%	fairness
4.	20.16%	equal amount for everyone	19.85%	equal choice
5.	16.13%	afraid to be punished	17.56%	punishment brings no advantage

Table notes: Multiple answers are possible.

Turning to the recipients, we see that fairness also plays a role, but only in third place. More important are the cost of awarding punishment points and the distribution chosen (either unequal or equal). In addition, 17.56% state that punishment is not beneficial because it is costly. Taking all participants together who indicate that they act payoff-maximizing, never punish, punishment is costly, and punishment does not bring any benefit (categories 3, 7, 9, and 10 of the open question), we find that 48.85% of the answers correspond to at least one

[68] The exact wording in German: In diesem Experiment konnten Sie als Teilnehmer A über eine Verteilung abstimmen.Welche Überlegungen haben Sie dabei angestellt? / In diesem Experiment konnten Sie als Teilnehmer B Abzugspunkte verteilen. Welche Überlegungen haben Sie dabei angestellt?

of these points. Of those, almost 90% actually do not punish. Since this is consistent with the actual punishment decision, according to which 42.22% never punish, we can confirm that participants do indeed make these self-interested and payoff-maximizing considerations.

Comparing the top five responses of recipients between treatments in Table 7.17, (Table A.10 in Appendix A.4 displays answers separately by default and mechanism.) shows that the fairness motive is more pronounced in 'random-unequal' than in the other treatments, while the monetary argument (punishment costs and brings no benefit) seems to be more important in the other treatments, especially in 'grade-unequal' and 'grade-equal'.

Table 7.17.: Top-5 Answers of Recipients by Treatments

	Answer	Overall	Random-unequal	Random-equal	Grade-unequal	Grade-equal
1.	punishment costs / is expensive	28.24%	16.67%	32.26%	32.26%	33.33%
2.	unequal choice	23.66%	13.89%	32.26%	22.58%	27.27%
3.	fairness	21.37%	30.56%	22.58%	19.35%	12.12%
4.	equal choice	19.85%	22.22%	12.90%	19.35%	24.24%
5.	punishment brings no advantage	17.56%	11.11%	9.68%	19.35%	30.30%

Table notes: Multiple answers are possible.

Summary 7.9: Open Question

Several conclusions can be drawn from the answers to the open question. First, fairness and payoff maximization are relevant motives for decision-makers. Second, the punishment decision is influenced by the punishment costs and the actual choice of the decision-maker. Thus, the analysis of the stated preferences of recipients confirms the observed preferences. In particular, the regression analyses show that recipients are sensitive to the actual choice of the decision-maker. The choice and the implemented outcome are identified as the main motives for the punishment decision, which is confirmed by their answers to the open question.

Hurdle Model and Cluster Analysis

As a further robustness check, a hurdle model is estimated. The Double Hurdle approach, originally introduced by Cragg (1971), is extended by Engel and Moffatt (2014) to implement

it in Stata and make it applicable to experimental panel data. The advantage of the Double Hurdle model is that it allows the combined estimation of two different processes: first, the decision to punish, and second, the amount of punishment for those who punish. Thus, the first hurdle can be interpreted as a probability model that captures the effect of different punishment motives on the probability of punishment. In addition, the second hurdle is a censored tobit model that determines the amount of assigned punishment for those participants who passed the first hurdle and thus decided to punish. Since many participants do not punish, there are many zeros for the punishment variable. Consequently, a Double Hurdle model may provide new and meaningful insights into what drives punishment behavior.

For the estimation, the three most relevant punishment motives, *choice*, *initiation*, and *pivotality*, are used as hurdle variables. The goal is to estimate their influence on the probability of passing the first hurdle and thus being punished. Furthermore, each model is estimated with and without additional control variables. However, since neither treatment effects nor further insights emerged, we do not report the estimation results.

In addition, we examine individual decisions using cluster analysis to move away from average decisions across participants. Therefore, hierarchical clustering using the ward linkage method is used to identify different punishment types of participants. However, since this analysis does not provide new insights or lead to other types of punishment than the ones already identified (choice, initiator, pivotality), the results of the cluster analysis are not discussed further here.

7.4.6. Comparison to Bartling et al. (2015)

Since the experimental design is adapted from Bartling et al. (2015), this section compares the results of our study to Bartling et al. (2015). First, we discuss the punishment points assigned for an unequal outcome. Second, an econometric comparison of the different punishment motives is presented. In order to compare the overall effects of the different punishment motives and how people assign responsibility, the aggregated data of this study is used. Thus, the four treatments are not separated.

Responsibility Attribution for an Unequal Outcome

Table 7.18 presents the average punishment points for each decision-maker in each of the eight possible decision sequences for both studies. Due to methodological differences[69] in the experimental design, the data structure is different, so different tests of significance are

[69]We use the direct response method while Bartling et al. (2015) use the strategy method.

applied. Bartling et al. (2015) report statistical differences using a Wilcoxon signed rank test. In this experiment, however, significant differences are calculated with a Wilcoxon rank sum test because the observations are independent.

Not surprisingly, the punishment points assigned are lower for an equal choice than for an unequal choice. Since the unequal outcome and the corresponding punishment points are of primary interest, the focus is on the upper part of Table 7.18, where the unequal allocation is implemented. Here, four categories of decision-makers are discernible, as already classified in Section 7.4.1.[70]

As already addressed in Section 7.4.2, the main result of Bartling et al. (2015), that the pivotal decision-maker is punished significantly more than the other decision-makers, is not confirmed in this study. Instead, the *initiator*, i.e., the first decision-maker who chooses the unequal allocation[71], is punished significantly more than the pivotal decision-maker. For example, a close look at decision sequence four (e-u-u) illustrates this difference. In the experiment by Bartling et al. (2015), the second decision-maker, the initiator, receives an average of 1.83 punishment points and the third pivotal decision-maker receives 2.33 points ($p < 0.01$). In contrast, in our study, the initiator receives 2.64 punishment points and the pivotal decision-maker 1.58 points, which is also highly significant at the 1%-level, but in the opposite direction. However, the punishment differences between the two studies are only visible in sequences three and four. In sequences one and two (u-u-u and u-u-e), the initial and pivotal decision-makers receive almost the same amount of punishment points in both studies.

In the theoretical and empirical literature, delegating a decision to avoid punishment has been found to be effective (among others, see e.g. Bartling and Fischbacher 2012; Oexl and Grossman 2012). In our setting, the second decision-maker who chooses equal when the first decision-maker chooses unequal (sequence 3 and 5) can be seen as delegating the final decision to the third decision-maker.[72] As predicted by the empirical results, this decision-maker is punished significantly less than the third decision-maker who chooses the unequal outcome, which is then implemented for the whole group (Sequence 3, decision-maker 2: 0.13 vs. decision-maker 3: 2.01, $p < 0.001$). As a similar pattern is found in Bartling et al. (2015) delegation is effective in this setting.

[70] The original classification by Bartling et al. (2015) differs in some respects, as one category is named differently. However, to obtain a clear delineation of the different punishment motives, the categories are applied as discussed in Section 7.4.1.

[71] In the wording of Bartling et al. (2015) the decision-maker with an unkind intention

[72] One could argue that all decision-makers in the first position who choose equal are postponing the decision of how to start the sequence. However, since they can also be classified as having a kind intention, the distinction between delegating and merely acting kindly is more difficult to make.

Table 7.18.: Comparison to Bartling et al. (2015)

Allocation	Decision sequence	Decision-maker 1	Decision-maker 2	Decision-maker 3	Pivotal vs. all	Pivotal vs. initiator	Initiator vs. all
		Bartling et al. (2015)			**Wilkoxon sign rank**		
Unequal	u-u-u	1.50	1.85	0.86	p<0.001	p=0.294	
	u-u-e	1.86	1.92	0.26	p<0.001	p=0.960	
	u-e-u	1.68	0.07	2.39	p<0.001	p=0.006	
	e-u-u	0.11	1.83	2.33	p<0.001	p=0.012	
Equal	u-e-e	1.33	0.10	0.08			
	e-u-e	0.17	1.43	0.08			
	e-e-u	0.06	0.03	0.92			
	e-e-e	0.08	0.07	0.03			
		This study			**Wilkoxon ranksum**		
Unequal	u-u-u	1.33	1.00	0.87	p=0.946	p=0.7645	p=0.628
	u-u-e	2.00	2.02	0.09	p=0.002	p=0.782	p<0.001
	u-e-u	2.67	0.13	2.01	p=0.025	p=0.201	p<0.001
	e-u-u	0.00	2.64	1.58	p=0.129	p=0.009	p<0.001
Equal	u-e-e	1.68	0.09	0.12			
	e-u-e	0.05	1.41	0.05			
	e-e-u	0.03	0.08	2.26			
	e-e-e	0.18	0.13	0.17			

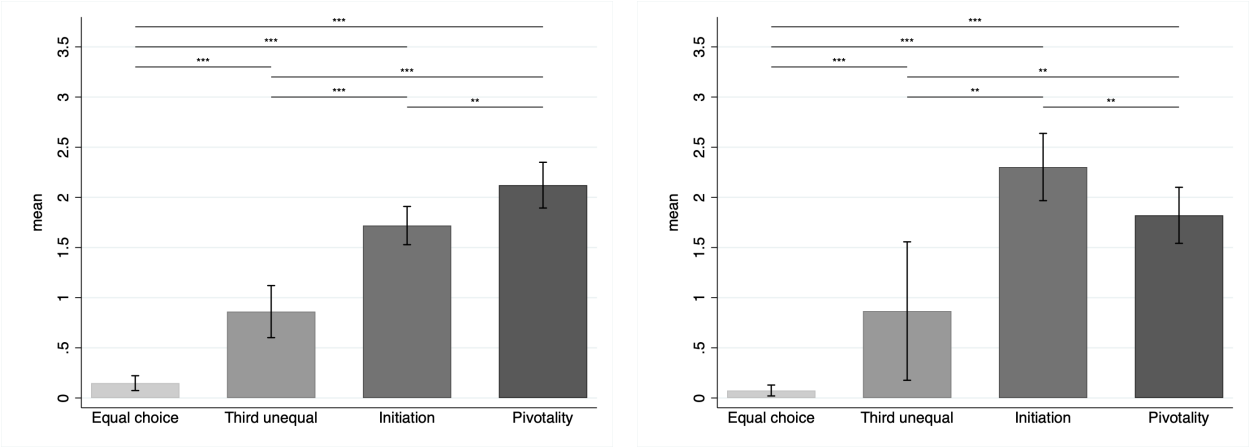
Table notes: "u" denotes a choice of the unequal allocation; "e" denotes a choice of the equal allocation. The three rightmost columns show p-values of Wilcoxon signed rank tests for Bartling et al. (2015) and of ranksum test for this study comparing the punishment for the pivotal decision-maker to the punishment for the two other decision-makers ("pivotal vs. all") and to the punishment for the initiator only ("pivotal vs. initiator"). Additionally, in this study a comparison between the punishment for the initiator to the punishment for the two other decision-makers ("initiator vs. all") is added.

One might argue that this dataset already contains the treatment variation, or rather the induced status differences, that might have influenced the punishment decision. Therefore, as a robustness check, the same analysis is presented in Appendix A.4 with a restricted dataset (only the two random treatments), which leads to the same punishment behavior.

To further illustrate the punishment behavior in both studies, Figure 7.4 presents the average punishment points for unequal outcomes. It also indicates significant differences between categories of decision-makers. Overlapping categories, such as unequal choice or

unkindness, are excluded to provide a clear and exclusive delineation of decision-makers.

The left side of Figure 7.4 illustrates the average punishment in the experiment by Bartling et al. (2015), while our results are shown on the right side. As described earlier, the fundamental difference between the two experiments concerns the punishment of the initiator and the pivotal decision-maker. As it can be seen in Figure 7.4, the mean punishment for the pivotal decision-maker in the experiment of Bartling et al. (2015) is significantly higher than for the initiator. In contrast, in this study it is the other way around. The initiator receives the most punishment points (mean of 2.30 points), which is also higher than the average punishment of the pivotal decision-maker in the setting of Bartling et al. (2015) (mean of 2.12 points). The other two categories indicate that the punishment for an equal choice is relatively small and does not differ between the two experiments. The same is true for the third decision-maker, who chooses the unequal allocation when the other two have already decided on the unequal outcome. In both experiments, significant differences are found between all four categories of decision-makers, with the direction of these differences being opposite for initiation and pivotality.



(a) Bartling et al. (2015)

(b) This study

Notes: Mean of assigned punishment points (from 0 to 7) for the four different categories of decision-makers as defined in Section 7.4.1. The p-values in the left figure are from signrank test and in the right figure from ranksum test. Bars represent the 95% confidence interval. $^{*}p \leq 0.1$, $^{**}p \leq 0.05$,$^{***}p \leq 0.01$.

Figure 7.4.: Average Punishment If the Unequal Allocation Results

Econometric Comparison of Different Punishment Motives

In this section, an OLS regression is presented to explore punishment behavior in more detail and to examine the effects of different punishment motives. Therefore, we use the same

punishment variables as in Bartling et al. (2015) to directly compare the results. This study uses the entire dataset without distinguishing treatments.[73] In the beginning, in Models 1 through 8, each punishment motive is regressed separately on the assigned punishment points. Then, all motives are included together (Model 9). In addition, the last column of Table 7.19 presents the regression results of Bartling et al. (2015) for comparison with our study.[74]

Several conclusions can be drawn from the data in Table 7.19. First, each punishment motive has a significant positive effect on the punishment points assigned. The only exception is the equal choice, which has a negative effect. This is obvious because participants have only two choices, either equal or unequal. Second, each motive induces an increase of at least one punishment point. Being the initiator of an unequal outcome has the largest effect (Coef. 1.857). More precisely, the initiator receives on average 1.86 punishment points more than a decision-maker who is not the initiator. Since a decision-maker can only be pivotal in an unequal outcome, 'outcome X initiator' is added and specifies the initiator for the unequal outcome. Third, the highest explanatory power is found for an unequal (equal) choice, while an unkind intention has the second highest explanatory power. Since the two are closely related and differ only in the effect a choice has, similar explanatory power is expected (recall that the motive unkind means choosing unequally when the outcome is not fixed). Furthermore, we show that the explanatory power for initiation (and initiator for unequal outcome) is much higher than for pivotality (0.201 and 0.130 versus 0.069, respectively). In addition, the impact on assigned punishment points is also higher for initiation and initiator for the unequal outcome than for pivotality (Coef. 1.673 and 1.857 versus 1.312), although all are highly significant at the 1%-level.

[73] As in the previous section, the same regression with the random treatments is presented in the Appendix A.4.

[74] To compare our results and the R^2 with the experiment of Bartling et al. (2015), we use an OLS regression here. However, since a tobit regression fits our data better, the corresponding tobit regression can be found in Table A.12 in the Appendix A.4.

Table 7.19.: Comparison of Different Punishment Motives

Punishment Points	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	Bartling et al. (2015)
Outcome unequal	1.176***								-0.030	0.048
	(0.102)								(0.057)	(0.070)
Unequal choice		1.804***							1.096***	0.782***
		(0.153)							(0.160)	(0.137)
Unkind			1.756***							
			(0.131)							
Equal choice				-1.804***						
				(0.153)						
Initiator					1.673***				0.468*	0.517**
					(0.194)				(0.269)	(0.196)
Outcome X Initiator						1.857***			0.692***	0.289
						(0.211)			(0.235)	(0.232)
Outcome X Choice							1.802***			
							(0.148)			
Pivotal								1.312***	0.679***	0.403**
								(0.148)	(0.175)	(0.155)
Constant	0.282***	0.127***	0.177***	1.931***	0.277***	0.393***	0.265***	0.429***	0.128***	0.083**
	(0.039)	(0.033)	(0.037)	(0.161)	(0.035)	(0.038)	(0.037)	(0.041)	(0.036)	(0.037)
R^2-overall	0.117	0.297	0.260	0.297	0.201	0.130	0.205	0.069	0.309	0.281
N	2430	2430	2430	2430	2430	2430	2430	2430	2430	1728

Table notes: Random effects OLS regression with punishment points for decision-makers as dependent variable (from 0 to 7). All punishment variables are dummy variables. Outcome unequal equals 1 if the unequal allocation is implemented. Unequal choice equals 1 if the unequal allocation is chosen. Unkind equals 1 if a decision-maker opts for the unequal allocation. Initiation equals 1 if the decision-maker is the first one choosing unequal. Pivotal equals 1 if a decision-maker is pivotal for the unequal outcome. Standard errors are in parentheses and are clustered on 12 session in this study and 72 individuals in Bartling et al. (2015). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

In Model 9, all punishment motives are included simultaneously to estimate their joint effect on the assigned punishment points. Since equal and unequal choices are interchangeable, only unequal choices are considered. In addition, to have a cleaner comparison to Bartling et al. (2015), the interaction 'outcome X choice' and the motive unkindness[75] are not included in this model.

A comparison to Bartling et al. (2015) (Model 8 with 9) shows that an unequal choice is highly significant in both studies. However, the impact is slightly higher in this study, increasing punishment by about 1.10 points compared to 0.78 points in Bartling et al. (2015). In addition, the comparison reveals an important pattern between the initial and pivotal decision-makers. Not surprisingly, both punishment motives significantly affect the punishment points assigned. However, in this study, being the initiator is only weakly significant (10%-level). In contrast, being the initiator of the unequal outcome has a highly significant positive effect, leading to an average increase of 0.69 punishment points. In the experiment by Bartling et al. (2015), the initiator receives on average 0.52 points more, while the initiator of an unequal outcome does not significantly influence the assigned punishment points. The coefficient of the pivotal voter is also slightly different (0.68 vs. 0.40). Furthermore, the explanatory power in this study is higher than in the experiment of Bartling et al. (2015) (0.309 vs. 0.281). Since one can argue that the role assignment process in Bartling et al. (2015) is quite different from that in this study, the same OLS regression with only the two random treatments supports the described differences (see Table A.13 in Appendix A.4). When we take them together, we can confirm our previous result 5, according to which the initiator's influence and explanatory power is higher in our study.

As described in Section 2.1.3, Engl (2018) develops a theoretical model where he distinguishes two forms of responsibility (ex-ante and ex-post) to predict the punishment behavior. However, as it is based on the Structural Model Approach with pivotality as the main component, it is not directly applicable to this study.[76] Adapting the model to initiation as a measure of distance leads to a fundamental problem. The second (third) decision-maker can never be the initiator of the unequal outcome, since at least two decision-makers who choose unequally are required to implement the unequal outcome. Suppose the first decision-maker chooses equally and the second unequally; he is the initiator, while the third decision-maker can never be the initiator. Therefore, we will not discuss this aspect here.

[75] Even adding unkindness does not lead to significantly different results.

[76] Nevertheless, pivotality significantly affects the assigned punishment points in this study. Including ex-ante and ex-post responsibility, as formalized in Engl (2018), reveals a significant effect (Coef: 0.348, $p > 0.01$) of ex-ante responsibility with an explanatory power of 0.01, while no effect of ex-post is identified.

7.4.7. Actual Punishment versus Expected Punishment

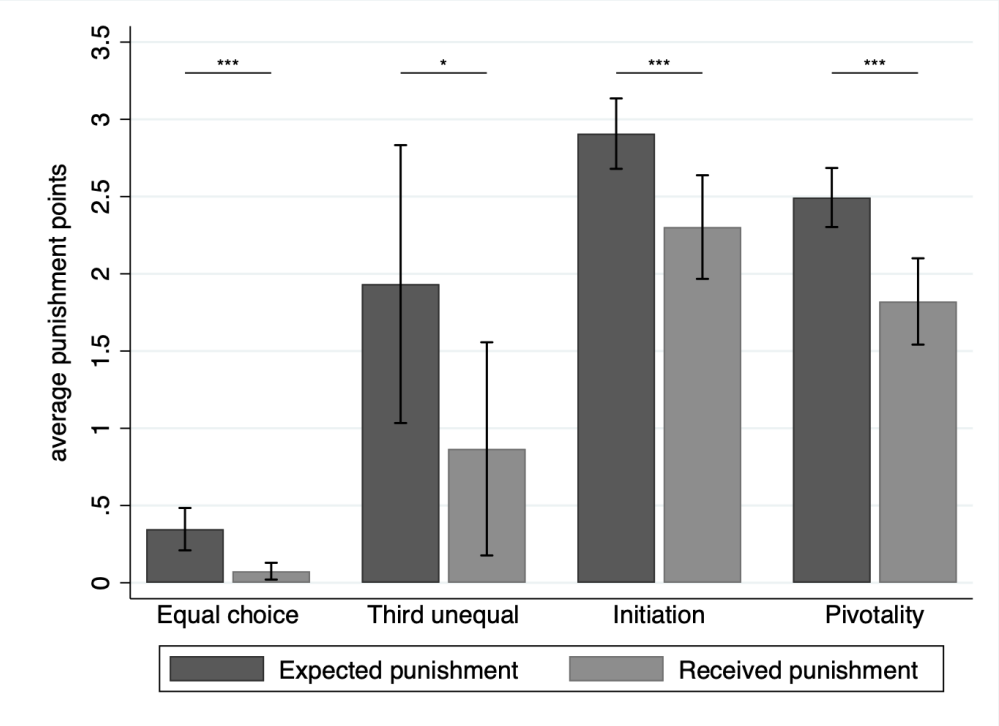
Finally, a brief comparison is given between the expected punishment and the punishment received. Since acts of omission are often used to avoid punishment, which has been shown to be effective in previous experiments (among others, see e.g. DeScioli et al. 2011b; Bartling and Fischbacher 2012) as well as in our setting, it is interesting to examine whether choosing the default is also expected to receive less punishment. More specifically, do decision-makers expect to be held less accountable if they choose the default or are assigned to their position through a legitimate process? To answer this question, we compare the decision-makers' expected punishment points with the punishment points they actually received.

Irrespective of the outcome, the decision-makers expect on average 0.77 punishment points, while the recipients receive 0.52 points. This difference is statistically significant at the 1%-level (t-test). Furthermore, a χ^2-test reveals differences between expected and actual punishment points ($\chi^2 = 84.509$, $p < 0.01$). For the unequal distribution, the difference between expected and received punishment points is 0.55 points (1.97 to 1.42) and is statistically significant (t-test, $p < 0.01$).

Figure 7.5 compares expected and actual punishment for the four main punishment motives. The most obvious difference is that decision-makers expect far more punishment points than they actually receive. Furthermore, with the exception of 'third unequal', these differences are highly significant at the 1%-level. Consequently, there are large differences between perceived responsibility (or expected assigned responsibility) and the self-assessment of responsibility by recipients. For example, Leibbrandt et al. (2012) find similar results in ten simple allocation games where a first player has to choose between two allocations and a second player can punish afterwards. Stable across allocations, the expected punishment is (almost) always higher in absolute terms than the expected reward. Unfortunately, the authors do not report significant differences between actual and received punishment (p. 759).

Tobit regressions are estimated to examine treatment differences in expected punishment.[77] The aim is to investigate whether decision-makers expect less (or more) punishment for an unequal (equal) default and whether the legitimacy of the group-building mechanism influences their expectations. Table A.14 in Appendix A.4 presents these results, from which several conclusions can be drawn. The default has a significant effect on the expected punishment. Thus, when the equal allocation is preselected, decision-makers expect on average 0.27 more punishment points than with an unequal default ($p < 0.01$). When we turn to

[77] A corresponding probit regression is also estimated, with similar results. It is therefore not discussed here.



Notes: Mean of assigned punishment points (from 0 to 7). The p-values are from two-tailed t-tests of treatment differences, assuming equal variances. Bars represent the 95% confidence interval. $^{*}p \leq 0.1$, $^{**}p \leq 0.05$,$^{***}p \leq 0.01$.

Figure 7.5.: Expected versus Actual Punishment if the Unequal Allocation Results

the treatment distinction (Model 3), an interesting aspect emerges. Compared to the baseline category random-unequal, decision-makers in grade-unequal expect significantly fewer penalty points (Coef. -0.26, $p < 0.1$), which is also found when socio-demographic control variables are added (Model 5). This should be treated with caution, as the treatment effect disappears when we add the punishment motive and may be due to the interaction between default and mechanism. The addition of the default choice dummy in Model 4 reveals that decision-makers expect to be punished significantly less if they choose the default option (Coef. -0.32, $p < 0.01$), which is stable across models and consistent with the actual behavior of decision-makers, as almost 50% of the choices are default choices. Turning to the last two Models 6 and 7, where punishment motives are included, leads to unexpected results. In terms of the actual punishment, an unequal choice is expected to be punished significantly more than an equal choice. However, being the initiator and being pivotal affect the expected punishment in the opposite direction. More specifically, the expected punishment for being the initiator is 0.57 points *lower* than for a decision-maker who is not the initiator ($p < 0.01$). Similar results are found for the pivotal decision-maker (Coef. -0.98, $p < 0.01$). To further

investigate the cause of these unexpected results, we split the dataset and consider only unequal outcomes. When the implemented outcome is the unequal allocation, the direction of these effects changes and is as expected, as initiation and pivotality positively affect the expected punishment.

Summary 7.10: Results of Expected Punishment

In addition to the general finding that decision-makers expect more punishment points than recipients, the tobit regression shows that decision-makers expect less punishment when they choose the default allocation. In addition, unequal choice is expected to lead to a large increase in assigned punishment points, higher than the actual increase by recipients. The main finding of actual punishment behavior, where the initiator is punished significantly more, is only observed for unequal outcomes. Taken together, however, the results on expected punishment behavior reinforce our conclusions from the previous sections.

8. Conclusion

Our research aims to answer the question how people assign individual responsibility for sequential group decisions with status differences and a preselected allocation, thus complementing similar studies by Bartling et al. (2015) and Duch et al. (2014). Therefore, we consider two dimensions: first, the way in which choices are made, either actively or passively by implementing a default option. Second, the influence of a more (less) legitimate group-building mechanism on attributed responsibility, given that the legitimacy of status differences affects the acceptance of inequalities.

From a theoretical perspective, we identify two approaches that build on the concept of causality and specify how responsibility should be attributed: either the causality of the decision-maker or the causality of the action. In this respect, the attribution of responsibility is mitigated or enhanced by various factors, such as intention (Alicke 2000; Malle et al. 2014) or being pivotal (Chockler and Halpern 2004; Engl 2018). The (experimental) literature confirms and extends these aspects. Acts of omission are punished less than acts of commission (Spranca et al. 1991; Vaal 1996). The default option is chosen more often and punished less because it is perceived as more appropriate or as a recommendation (McKenzie et al. 2006; Dhingra et al. 2012). Social norms and preferences, as well as various fairness and distributional criteria, specify how payoffs should be distributed (Bernhard et al. 2006; Cappelen et al. 2007), with procedural fairness being a key driver of fairness judgments (Bolton et al. 2005; Ku and Salmon 2013). Experiments such as those by Bartling et al. (2015), Anselm et al. (2022), and Duch et al. (2014), which analyze (sequential) group decisions, identify pivotality and proposal power as cues for assigning responsibility for unequal outcomes. However, to encourage acceptance of unpleasant outcomes, the process that leads to those outcomes and produces status differences should be fair and legitimate (Olson and Hafer 2001; Ridgeway 2001).

Thus, the first part of this study specifies the situational context, the decision maker's dispositional factors, and his intention to determine the assigned responsibility. It also emphasizes the heterogeneity of fairness and distributional preferences, so that no single criterion is (always) applicable. Consequently, different aspects shape and influence the assignment of responsibility, while counterfactual thinking takes place to compare the just assignment with the actual assignment.

However, before conducting the laboratory experiment, we carry out a pretest to examine

which two mechanisms (out of four) are the most and least legitimate. The pretest shows that a random group-building mechanism is the most legitimate, while the final grade is the least legitimate. In our experiment, these two mechanisms are then utilized to create (legitimate) status differences between decision-makers and recipients. The experimental design is adapted from Bartling et al. (2015) and extended by the two dimensions described above, resulting in a 2x2 design with either an equal or unequal allocation as default and a more (less) legitimate group-building mechanism (random or grade).

We analyze assigned responsibility as measured by the assignment of punishment points along two dimensions. First, at the extensive margin, we focus on the probability of punishment regardless of the amount of assigned punishment points. Therefore, a yes-no responsibility distinction is considered, consistent with the theoretical models of Shaver (1985) and Schultz et al. (1981). Second, and more importantly, we are interested in the degree of responsibility assigned, as we expected the above factors to mitigate or increase responsibility (theoretically proposed, i.e., by Spellman (1997) and Chockler and Halpern (2004)). Therefore, at the intensive margin, we estimate factors that influence the exact amount of assigned punishment points.

Our regression analyses highlight three dimensions on which responsibility is assigned: (1) An unequal choice significantly increases the likelihood of punishment and the punishment points awarded, having the largest effect and the strongest explanatory power. (2) Being the initiator and pivotal to the unequal outcome determines punishment behavior, with the initiator's influence being stronger. (3) Choosing the default leads to lower punishment as long as punishment motives are not included. Surprisingly, we find no significant treatment differences in punishment behavior when punishment motives are included, suggesting that punishment behavior is determined by punishment motives rather than the situational context.

In addition, we find a lower probability of punishment for women, who also assign fewer punishment points. Since we do not have a balanced sample, perhaps treatment effects would be more pronounced with more men. Further research should shed light on general gender differences in punishment behavior, and whether women generally punish less frequently and less intensely than men.

The most surprising finding is that we cannot confirm the pivotality effect found by Bartling et al. (2015) and Anselm et al. (2022). On the contrary, the initiator, i.e, the first decision-maker to choose the implemented (unequal) allocation, is held more responsible than the other decision-maker(s). In our probit and tobit regressions, being the initiator has a higher impact on the decision to punish and the number of punishment points awarded,

which is supported by our robustness checks. Interestingly, combining being the initiator with (not) choosing the default option shows that the initiator is punished significantly more regardless of the default option, while even more punishment points are assigned for passively choosing the unequal allocation (since it is the default). Conversely, choosing the default allocation and being the pivotal decision-maker can prevent from punishment or reduce the amount of punishment points assigned.

Our experiment differs from previous studies by Bartling et al. (2015) and Anselm et al. (2022) in two important ways. First, we include our treatment variations insofar as we induce status differences and include a default option. Second, we use a different elicitation method. The latter factor should not make a difference, as argued in Section 6.4. The strategy and direct response methods should induce the same behavior, so this cannot be the main reason for such differences. However, Falk et al. (2005) specify that the probability of punishment is the same for both methods, while the direct response method yields to significantly more punishment points. In addition, using the same experimental design, once with the strategy method and once with the direct response method, they demonstrate that behavior could differ (Cox and Deck 2005; Falk et al. 2003). Relatedly, even negligible variations in the decision context can significantly alter decisions and the resulting outcomes (Krupka and Weber 2013). As Konow (2000) notes, "most peoples' values on fairness may be accounted for by several fairly simple principles, any of which may dominate depending on the context. [...] People may weight justice principles differently or perceive and evaluate the factors relevant to even a single principle differently" (p. 1074), which may account for our results.

Previous theoretical and experimental research has mainly focused on the pivotal decision-maker as the focal point for attributing responsibility (among others, see e.g. Chockler and Halpern 2004; Bartling et al. 2015; Engl 2018). However, Spellman's crediting causality hypothesis (1997) presents a different view that could explain the punishment behavior in our study. It posits that a chain of events determines causality and distinguishes two chains, either temporal or causal. In causal chains, the first event is seen as more causal. Applying this to our study and assuming a 'chain of decision-makers', the first decision-maker would be more causal for the outcome and bear more responsibility than the subsequent decision-maker(s). Since this is exactly what we exhibit in our study, it is logical to punish the initiator who initiates the (unequal) decision sequence.

Although we do not find significant treatment differences, at least three important patterns emerge. First, we shed light on the importance of replications and extensions in (experimental) economics. Second, we reveal that a default may affect the punishment decision on some dimensions (such as pivotality), but to a lesser extent than theoretically predicted. Third,

in our study, induced status differences have no effect on assigned responsibility, which is completely unexpected. There are several reasons to explain this.

If an unpleasant outcome occurs, the cause and personal control are relevant in distributive situations. In this study, it is questionable whether participants consider the final grade to be under (full) personal control. If the final grade is controllable, individuals must take full responsibility and do not deserve compensation for negative consequences (Becker 2012; Cappelen et al. 2010). In addition, a person with a grade near or slightly above the median will be assigned the role of decision-maker, since the majority is worse in their session. However, if the majority is slightly better than him, he would be in the role of recipient. This may be due to pure luck, i.e., brute luck, since no one can influence this classification, and the composition of the sessions is completely random. However, the argument that it is due to controllable option luck, since more effort at school leading to a better grade would put him in a higher position, is also applicable.

Relatedly, it is unclear whether participants perceive the process by which roles are assigned to be fair or legitimate. If randomization or the final grade is perceived as a more (less) legitimate process, the resulting inequalities are more likely to be accepted and seen as less unfair. This may explain why recipients do not punish tolerating unequal payoff distributions. Alternatively, a decision-maker may be identified as a legitimate actor in a normative sense, which does not automatically imply empirical legitimacy and thus acceptance by the participants. Since the final grade as a group-building mechanism is not determined by the participants but by the experimenter, legitimacy in an empirical sense is not necessarily given, as argued in Section 4.2.2.

On the other hand, it is not just about the empirical legitimacy of the grade as such, but whether the classification according to the grade is empirically legitimate. In addition, the pretest and the end-of-experiment evaluation revealed differences in perceived legitimacy between the two procedures. Thus, they may not be large or relevant enough to influence punishment behavior in our setting. Alternatively, the final grade may be perceived as less legitimate than a random procedure, but this does not necessarily mean that a role assignment based on a random procedure is also legitimate. This is supported by the fact that we often find differences in absolute terms, i.e., the mean punishment for the initiator and the pivotal decision-maker (see Table 7.9), which are not significantly different.

Another explanation has to do with gender differences. Previous research has identified behavioral differences between men and women (Singer et al. 2006; Burnham 2018), which were confirmed in our pretest and experiment. Women rated both mechanisms rather more extreme than men, while at the same time the probability of punishment and the amount of

punishment points assigned are lower for them. Since our sample is not gender balanced, it is possible that differences in punishment are driven by gender differences, as we found mutually significant treatment differences when considering only men (see Table A.7, column 7). Consequently, significant treatment differences, and thus differences between group-building mechanisms, may be more pronounced with a more balanced sample.

Moreover, it is not obvious what a higher grade reveals. In addition, it is questionable whether the school grade is an equal opportunity for everyone and the process of achieving the final grade is fair (or is perceived to be fair). Although the school system in Germany is open to everyone and everyone can try to get the 'Abitur' and go to university, in reality the situation might be quite different. As Weiner (1993) points out, there are extenuating factors that explain why someone shows less effort and gets bad grades at school, i.e., a sick family member to care for. In such a case, it is perfectly acceptable for someone to not use his abilities to the fullest, so his school grade may not be as good as it otherwise would have been. However, since this is not something that the judging person (here, the recipient) can determine, he may not want to punish someone for a fact that he cannot fully control.

In this regard, Mikula (2003) emphasizes that the mere violation of an entitlement is not enough to hold someone responsible. Other mitigating factors, such as control over and intent behind the action, must be considered. Another explanation is provided by the SCT, since a status characteristic must be activated to be effective. Even if the activating conditions are not fully known or clear, the final grade is probably not a status characteristic. It might not define an assumption or belief about the skills, knowledge, or abilities of the decision-maker (Berger et al. 1972, pp. 244-245). Therefore, it would not determine the punishment behavior of recipients, who might perceive both status positions as almost random.

However, even if one criticizes the use of the final grade as a (less) legitimate group-building mechanism, it is in fact relevant in real life and represents the most important basis for the allocation of study places or entry positions into the job market (Weiss and Fershtman 1998; Brandstätter and Farthofer 2002). In addition, independent of the subjective influence of teachers, schools, and social background (Ingenkamp 1997; Süß 2001), the final grade represents in part the effort during a two-year school period and thus provides information about the person (Camara 2005).

Although the right to decide is usually perceived as an intrinsic value in itself (Bartling et al. 2014b), it is also possible that some participants do not really want to decide how to allocate the endowment, especially if a reciprocal reaction of the recipients is expected. Consistent with the experimental findings on delegation (Bartling and Fischbacher 2012; Oexl and Grossman 2012) and (willful) ignorance (Conrads and Irlenbusch 2013; Wieland 2016),

people do not always want to decide and (sometimes) prefer to shift the responsibility for the outcome to someone else. Since this may be anticipated or perceived similarly by recipients who also do not want to decide, an accommodating reaction by recipients is conceivable.

Even if recipients perceive decision-makers as responsible and think they should punish them, there are several reasons why they may not assign punishment points. The most important reason may be their self-interest and payoff maximization, which is predicted by rational choice theory and is one of the main reasons mentioned in the open question. In addition, recipients may perceive harm in punishment, so they avoid it by not punishing.

In addition to the punishing behavior of the recipients, we also found some interesting patterns in the choices of the decision-makers. First, far more decision-makers choose the equal allocation in our study than in the experiment of Bartling et al. (2015), while several factors could explain this behavior. For example, since more women participated in our experiment, and women are more inequality averse compared to men (Essen and Ranehill 2011), this could explain the high number of equal choices. Alternatively, a fair or equal distribution might be considered more appropriate because the endowment is given for free, like a 'windfall money' (Cappelen et al. 2007). Since one allocation is always preselected, this might have influenced behavior. Interestingly, the default has an unexpected effect on decision-makers' choices. One might have expected more selfish choices with an unequal default, since choosing the default is (usually) more accepted, and people often keep the default (Dhingra et al. 2012). Of course, this could be an artifact of the experimental design, since the group-building mechanism or the group decision could have induced such behavior. However, future research should focus on equal and unequal default choices in group decisions to investigate whether this behavior is consistent and explainable. A more systematic approach is warranted to fully understand this unexpected finding.

Of course, our design is not free from criticism. We use a fixed set of choices with only two allocations, which restricts the participants' decision process. The unequal allocation is quite extreme, encouraging equal choices because the unequal distribution may be seen as too unequal. It is possible that an intermediate allocation would result in more choice diversity and more (or less) punishment. Perhaps different punishment behavior could be achieved with a binary punishment option that assigns a fixed number of punishment points (or no punishment at all), as used in Molenmaker et al. (2016). A fixed set of choices could be perceived as the experimenter's recommendation, especially with the preselected option. As a result, true intentions could not be discerned. Duch and Stevenson (2014) extended the set of choices to three allocations, where the third allocation was also unequal, but not as unequal as the original 9-1 allocation. This could be a way to still have unequal choices, but

to a lesser extent. A different experimental design with a greater variety of allocations would likely reveal different intentions and choices by decision-makers and recipients.

Since the forces that determine assigned responsibility are not fully understood, future research should shed light on other relevant factors, dimensions, and (or) aspects that might mitigate or enhance assigned responsibility. Not to mention the comparison between expected and actual punishment. As already addressed in the empirical literature, delegation and avoidance reduce the assigned punishment, although this does not automatically mean that decision-makers feel less responsible. Further research on how responsible parties perceive themselves and how they are held responsible by affected parties and uninvolved third parties would be an interesting area for future research.

In addition, as shown in the experiments of Cappelen et al. (2007, 2010), people may value fairness considerations differently when the endowment is earned in a production phase. However, even when participants do not earn their endowment, using a different procedure, such as the slider task, may reveal different drivers for the punishment decision. Explicitly stating how many sliders each participant solved, so that it is apparent how much effort each participant put into the task, might elicit a completely different punishment behavior. Furthermore, the definition of status specifies that a ranking in a socially recognized hierarchy is required to assign status differences. Since the exact ranking is not publicly announced in our study, it is possible that participants do not believe in status differences or do not perceive them as real. Future research should focus on more overtly declaring status differences, such as publicly announcing the ranking or explicitly stating that one group has higher status, i.e., by assigning stars as in Ball et al. (2001).

Finally, we hope to have provided a better understanding of how people assign responsibility in sequential group decisions. The topic is of widespread importance affecting many aspects of everyday life, so research should not stop here. Perhaps we have been able to provide a new perspective on people's focus on holding others accountable, while leaving room for avenues for further research.

Bibliography

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond (2019). "Preferences for Truth-Telling". In: *Econometrica* 87.4, pp. 1115–1153.

Adams, J. Stacy (1963). "Towards an Understanding of Inequity." In: *The Journal of Abnormal and Social Psychology* 67.5, pp. 422–436.

Adriaans, Jule, Philipp Eisnecker, and Stefan Liebig (2019). "A Comparison of Earnings Justice throughout Europe: Widespread Approval in Germany for Income Distribution According to Need and Equity". In: *DIW Weekly Report*.

Agranov, Marina and Pietro Ortoleva (2017). "Stochastic Choice and Preferences for Randomization". In: *Journal of Political Economy* 125.1, pp. 40–68.

– (2022). "Revealed Preferences for Randomization: An Overview". In: *AEA Papers and Proceedings* 112, pp. 426–430.

Akaike, H. (1974). "A New Look at the Statistical Model Identification". In: *IEEE Transactions on Automatic Control* 19.6, pp. 716–723.

Akbaş, Merve, Dan Ariely, and Sevgi Yuksel (2019). "When is Inequality Fair? An Experiment on the Effect of Procedural Justice and Agency". In: *Journal of Economic Behavior & Organization* 161, pp. 114–127.

Akerlof, George A. and Rachel E. Kranton (2000). "Economics and Identity". In: *Quarterly Journal of Economics* 115.3, pp. 715–753.

Albrecht, Konstanze, Emma von Essen, Klaus Fliessbach, and Armin Falk (2013). "The Influence of Status on Satisfaction with Relative Rewards". In: *Frontiers in Psychology* 4.

Alicke, Mark D. (2000). "Culpable Control and the Psychology of Blame." In: *Psychological Bulletin* 126.4, pp. 556–574.

Anderson, Christopher (1995). *Blaming the Government: Citizens and the Economy in Five European Democracies*. M.E. Sharpe Inc.

Anderson, Christopher J. (2003). "The Psychology of Doing Nothing: Forms of Decision Avoidance Result from Reason and Emotion." In: *Psychological Bulletin* 129.1, pp. 139–167.

Andreoni, James and B. Douglas Bernheim (2009). "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects". In: *Econometrica* 77.5, pp. 1607–1636.

Andreoni, James and John Miller (2002). "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism". In: *Econometrica* 70.2, pp. 737–753.

Andreoni, James and Lise Vesterlund (2001). "Which is the Fair Sex? Gender Differences in Altruism". In: *The Quarterly Journal of Economics* 116.1, pp. 293–312.

Anselm, Regina, Deepti Bhatia, Urs Fischbacher, and Jan Hausfeld (2022). "Blame and Praise: Responsibility Attribution Patterns in Decision Chains". In: *Research Paper Series Thurgau Institute of Economics and Department of Economics at the University of Konstanz*, pp. 1–44.

Arneson, Richard J. (1989). "Equality and Equal Opportunity for Welfare". In: *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 56.1, pp. 77–93.

Aytaç, Selim Erdem (2018). "Relative Economic Performance and the Incumbent Vote: A Reference Point Theory". In: *The Journal of Politics* 80.1, pp. 16–29.

Bagwell, Laurie Simon and B. Douglas Bernheim (1996). "Veblen Effects in a Theory of Conspicious Consumption". In: *American Economic Review* 86.3, pp. 349–373.

Balafoutas, Loukas, Nikos Nikiforakis, and Bettina Rockenbach (2014). "Direct and Indirect Punishment Among Strangers in the Field". In: *Proceedings of the National Academy of Sciences* 111.45, pp. 15924–15927.

Ball, Sheryl and Catherine C. Eckel (1998). "The Economic Value of Status". In: *Journal of Socio-Economics* 27.4, pp. 495–514.

Ball, Sheryl, Catherine C. Eckel, Philip J. Grossman, and William Zame (2001). "Status in Markets". In: *The Quarterly Journal of Economics* 116.1, pp. 161–188.

Banzhaf, John F. (1964). "Weighted Voting Doesn't Work: A Mathematical Analysis". In: *Rutgers Law Review* 19, p. 317.

Bardsley, Nicholas (2008). "Dictator Game Giving: Altruism or Artefact?" In: *Experimental Economics* 11.2, pp. 122–133.

Bartels, Daniel M. (2008). "Principled Moral Sentiment and the Flexibility of Moral Judgment and Decision Making". In: *Cognition* 108.2, pp. 381–417.

Bartling, Björn, Florian Engl, and Roberto A. Weber (2014a). "Does Willful Ignorance Deflect Punishment? – An Experimental Study". In: *European Economic Review* 70, pp. 512–524.

Bartling, Björn, Ernst Fehr, and Holger Herz (2014b). "The Intrinsic Value of Decision Rights". In: *Econometrica* 82.6, pp. 2005–2039.

Bartling, Björn, Ernst Fehr, and Yagiz Özdemir (2020). "Does Market Interaction Erode Moral Values?" In: *CESifo Working Paper* 8546.

Bartling, Björn and Urs Fischbacher (2012). "Shifting the Blame: On Delegation and Responsibility". In: *Review of Economic Studies* 79.1, pp. 67–87.

Bartling, Björn, Urs Fischbacher, and Simeon Schudy (2015). "Pivotality and Responsibility Attribution in Sequential Voting". In: *Journal of Public Economics* 128, pp. 133–139.

Battigalli, Pierpaolo and Martin Dufwenberg (2007). "Guilt in Games". In: *American Economic Review* 97.2, pp. 170–176.

Bauer, Alexander Max, Frauke Meyer, Jan Romann, Mark Siebel, and Stefan Traub (2022). "Need, Equity, and Accountability". In: *Social Choice and Welfare*.

Becker, Alice (2012). "Accountability and the Fairness Bias: the Effects of Effort vs. Luck". In: *Social Choice and Welfare* 41.3, pp. 685–699.

Bell, David E. (1982). "Regret in Decision Making under Uncertainty". In: *Operations Research* 30.5, pp. 961–981.

Bellemare, Charles, Alexander Sebald, and Sigrid Suetens (2017). "A Note on Testing Guilt Aversion". In: *Games and Economic Behavior* 102, pp. 233–239.

Bénabou, Roland and Jean Tirole (2006). "Incentives and Prosocial Behavior". In: *American Economic Review* 96.5, pp. 1652–1678.

Berger, Joseph, Bernard P. Cohen, and Morris Zelditch (1972). "Status Characteristics and Social Interaction". In: *American Sociological Review* 37.3, p. 241.

Berger, Joseph, M. Hamit Fisek, Robert Z. Norman, and Morris Zeldritch (1966). *Status Characteristics and Social Interaction. An expectation-states approach.*

Bernhard, Helen, Ernst Fehr, and Urs Fischbacher (2006). "Group Affiliation and Altruistic Norm Enforcement". In: *American Economic Review* 96.2, pp. 217–221.

Berry, James, Lucas C. Coffman, Douglas Hanley, Rania Gihleb, and Alistair J. Wilson (2017). "Assessing the Rate of Replication in Economics". In: *American Economic Review* 107.5, pp. 27–31.

Bettencourt, B. Ann, Kelly Charlton, Nancy Dorr, and Deborah L. Hume (2001). "Status Differences and In-Group Bias: A Meta-Analytic Examination of the Effects of Status Stability, Status Legitimacy, and Group Permeability." In: *Psychological Bulletin* 127.4, pp. 520–542.

Bicchieri, Cristina, Eugen Dimant, Simon Gächter, and Daniele Nosenzo (2022). "Social Proximity and the Erosion of Norm Compliance". In: *Games and Economic Behavior* 132, pp. 59–72.

Bilén, David, Anna Dreber, and Magnus Johannesson (2021). "Are Women More Generous than Men? A Meta-Analysis". In: *Journal of the Economic Science Association* 7.1, pp. 1–18.

Blader, Steven L. and Ya-Ru Chen (2014). "What's in a Name? Status, Power, and Other Forms of Social Hierarchy". In: *The Psychology of Social Status*. Springer New York, pp. 71–95.

Blount, Sally (1995). "When Social Outcomes aren't Fair: the Effect of Causal Attributions on Preferences". In: *Organizational Behavior and Human Decision Process* 63.2, pp. 131–144.

Blue, Philip R., Jie Hu, Xueying Wang, Eric van Dijk, and Xiaolin Zhou (2016). "When Do Low Status Individuals Accept Less? The Interaction between Self- and Other-Status during Resource Distribution". In: *Frontiers in Psychology* 7.

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch (2014). "hroot: Hamburg Registration and Organization Online Tool". In: *European Economic Review* 71, pp. 117–120.

Bolton, Gary E., Jordi Brandts, and Axel Ockenfels (2005). "Fair Procedures: Evidence from Games Involving Lotteries". In: *The Economic Journal* 115.506, pp. 1054–1076.

Bolton, Gary E. and Elena Katok (1995). "An Experimental Test for Gender Differences in Beneficent Behavior". In: *Economics Letters* 48.3-4, pp. 287–292.

Bolton, Gary E. and Axel Ockenfels (2000). "ERC: A Theory of Equity, Reciprocity, and Competition". In: *The American Economic Review* 90.1, pp. 166–193.

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

Bosman, Ronald, Matthias Sutter, and Frans van Winden (2005). "The Impact of Real Effort and Emotions in the Power-to-Take Game". In: *Journal of Economic Psychology* 26.3, pp. 407–429.

Bosman, Ronald and Frans Van Winden (2002). "Emotional Hazard in a Power-to-Take Experiment". In: *The Economic Journal* 112.476, pp. 147–169.

Brandstätter, Hermann and Alois Farthofer (2002). "Studienerfolgsprognose - Konfigurativ oder Linear Additiv?" In: *Zeitschrift für Differentielle und Diagnostische Psychologie* 23(4), pp. 381–391.

Brandts, Jordi and Gary Charness (2000). "Hot vs. Cold: Sequential Responses and Preference Stability in Experimental Games". In: *Experimental Economics* 2.3, pp. 227–238.

– (2003). "Truth or Consequences: An Experiment". In: *Management Science* 49.1, pp. 116–130.

– (2011). "The Strategy Versus the Direct-Response Method: a First Survey of Experimental Comparisons". In: *Experimental Economics* 14.3, pp. 375–398.

Brandts, Jordi and Carles Solà (2001). "Reference Points and Negative Reciprocity in Simple Sequential Games". In: *Games and Economic Behavior* 36.2, pp. 138–157.

Brewer, Marilynn B. and Roderick M. Kramer (1986). "Choice Behavior in Social Dilemmas: Effects of Social Identity, Group Size, and Decision Framing." In: *Journal of Personality and Social Psychology* 50.3, pp. 543–549.

Brock, J. Michelle, Andreas Lange, and Erkut Y. Ozbay (2013). "Dictating the Risk: Experimental Evidence on Giving in Risky Environments". In: *American Economic Review* 103.1, pp. 415–437.

Brosig, Jeannette, Joachim Weimann, and Chun-Lei Yang (2003). "The Hot Versus Cold Effect in a Simple Bargaining Experiment". In: *Experimental Economics* 6.1, pp. 75–90.

Brütt, Katharina, Arthur Schram, and Joep Sonnemans (2020). "Endogenous Group Formation and Responsibility Diffusion: An Experimental Study". In: *Games and Economic Behavior* 121, pp. 1–31.

Bundesamt, Destatis Statistisches (2022). *Absolventen und Abgänger: Deutschland, Schuljahr, Geschlecht, Schulart, Schulabschlüsse*. last retrieved 20.12.2022. URL: `https://www-genesis.destatis.de/genesis/online?operation=previous&levelindex=1&step=1&titel=Ergebnis&levelid=1671608705090&acceptscookies=false#abreadcrumb`.

Burnham, Terence C. (2018). "Gender, Punishment, and Cooperation: Men Hurt Others to Advance Their Interests". In: *Socius: Sociological Research for a Dynamic World* 4, p. 237802311774224.

Butler, Jeffrey V. (2014). "Trust, Truth, Status and Identity: An Experimental Inquiry". In: *The B.E. Journal of Theoretical Economics* 14.1, pp. 293–338.

Camara, Wayne J. (2005). "Broadening Predictors of College Success and the Impact of Cognitive Predictors". In: *Choosing students: Higher Education Admissions Tools for the 21st Century*. Ed. by Wayne J. Camara and Ernest W. Kimmel. Routledge, pp. 53–79.

Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan, Emma Heikensten, Felix Holzmeister, Taisuke Imai, Siri Isaksson, Gideon Nave, Thomas Pfeiffer, Michael Razen, and Hang Wu (2016). "Evaluating Replicability of Laboratory Experiments in Economics". In: *Science* 351.6280, pp. 1433–1436.

Campbell, Anne (2002). *A Mind Of Her Own*. Oxford University Press.

Campello, Daniela and Cesar Zucco (2016). "Presidential Success and the World Economy". In: *The Journal of Politics* 78.2, pp. 589–602.

Cappelen, Alexander W., Astri Drange Hole, Erik Ø. Sørensen, and Bertil Tungodden (2007). "The Pluralism of Fairness Ideals: An Experimental Approach". In: *American Economic Review* 97.3, pp. 818–827.

Cappelen, Alexander W., Ulrik H. Nielsen, Bertil Tungodden, Jean-Robert Tyran, and Erik Wengström (2015). "Fairness is Intuitive". In: *Experimental Economics* 19.4, pp. 727–740.

Cappelen, Alexander W., Erik Ø. Sørensen, and Bertil Tungodden (2010). "Responsibility for What? Fairness and Individual Responsibility". In: *European Economic Review* 54.3, pp. 429–441.

Carey, John M., Richard G. Niemi, and Lynda W. Powell (2000). "Incumbency and the Probability of Reelection in State Legislative Elections". In: *The Journal of Politics* 62.3, pp. 671–700.

Carpenter, Jeffrey and Emiliano Huet-Vaughn (2019). "Real-effort Tasks". In: *Handbook of Research Methods and Applications in Experimental Economics*. Edward Elgar Publishing Limited, pp. 368–383.

Carpenter, Jeffrey and Peter Hans Matthews (2009). "What Norms Trigger Punishment?" In: *Experimental Economics* 12.3, pp. 272–288.

Carroll, Gabriel D., James J. Choi, David Laibson, Brigitte C. Madrian, and Andrew Metrick (2009). "Optimal Defaults and Active Decisions". In: *Quarterly Journal of Economics* 124.4, pp. 1639–1674.

Cason, Timothy N., Lata Gangadharan, and Philip J. Grossman (2022). "Gender, Beliefs, and Coordination with Externalities". In: *Journal of Public Economics* 214, p. 104744.

Cason, Timothy N. and Vai-Lam Mui (1998). "Social Influence in the Sequential Dictator Game". In: *Journal of Mathematical Psychology* 42.2-3, pp. 248–265.

Çelen, Boğaçhan, Andrew Schotter, and Mariana Blanco (2017). "On Blame and Reciprocity: Theory and Experiments". In: *Journal of Economic Theory* 169, pp. 62–92.

Cettolin, Elena and Arno Riedl (2019). "Revealed Preferences under Uncertainty: Incomplete Preferences and Preferences for Randomization". In: *Journal of Economic Theory* 181, pp. 547–585.

Charness, Gary and Yan Chen (2020). "Social Identity, Group Behavior, and Teams". In: *Annual Review of Economics* 12.1, pp. 691–713.

Charness, Gary and Martin Dufwenberg (2006). "Promises and Partnership". In: *Econometrica* 74.6, pp. 1579–1601.

Charness, Gary and Uri Gneezy (2012). "Strong Evidence for Gender Differences in Risk Taking". In: *Journal of Economic Behavior & Organization* 83.1, pp. 50–58.

Chaudhuri, Ananish (2010). "Sustaining Cooperation in Laboratory Public Goods Experiments: a Selective Survey of the Literature". In: *Experimental Economics* 14.1, pp. 47–83.

Chen, Daniel L., Martin Schonger, and Chris Wickens (2016). "oTree—An Open-source Platform for Laboratory, Online, and Field Experiments". In: *Journal of Behavioral and Experimental Finance* 9, pp. 88–97.

Chen, Ya-Ru, Randall S. Peterson, Damon J. Phillips, Joel M. Podolny, and Cecilia L. Ridgeway (2012). "Introduction to the Special Issue: Bringing Status to the Table—Attaining, Maintaining, and Experiencing Status in Organizations and Markets". In: *Organization Science* 23.2, pp. 299–307.

Chen, Yan and Sherry Xin Li (2009). "Group Identity and Social Preferences". In: *American Economic Review* 99.1, pp. 431–457.

Cherry, Todd L., Peter Frykblom, and Jason F. Shogren (2002). "Hardnose the Dictator". In: *American Economic Review* 92.4, pp. 1218–1221.

Chockler, Hana and Joseph Y. Halpern (2004). "Responsibility and Blame: A Structural-Model Approach". In: *Journal of Artificial Intelligence Research* 22, pp. 93–115.

Claeskens, Gerda and Maarten Jansen (2015). "Model Selection and Model Averaging". In: *International Encyclopedia of the Social & Behavioral Sciences*. Elsevier, pp. 647–652.

Cohen, Gerald Allan (1989). "On the Currency of Egalitarian Justice". In: *Ethics* 99.4, pp. 906–944.

– (2011). *On the Currency of Egalitarian Justice, and other Essays in Political Philosophy*. Princeton University Press, Princeton.

Conrads, Julian and Bernd Irlenbusch (2013). "Strategic Ignorance in Ultimatum Bargaining". In: *Journal of Economic Behavior and Organization* 92, pp. 104–115.

Cox, Gary W. and Scott Morgenstern (1993). "The Increasing Advantage of Incumbency in the U. S. States". In: *Legislative Studies Quarterly* 18.4, p. 495.

Cox, James C. and Cary A. Deck (2005). "On the Nature of Reciproal Motives". In: *Economic Inquiry* 43.3, pp. 623–635.

Cox, James C. and Daniel T. Hall (2010). "Trust with Private and Common Property: Effects of Stronger Property Right Entitlements". In: *Games* 1.4, pp. 527–550.

Cox, James C., Maroš Servátka, and Radovan Vadovič (2016). "Status Quo Effects in Fairness Games: Reciprocal Responses to Acts of Commission versus Acts of Omission". In: *Experimental Economics* 20.1, pp. 1–18.

Cragg, John G. (1971). "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods". In: *Econometrica* 39.5, p. 829.

Crandall, Christian S. and Ryan K. Beasley (2001). "A Perceptual Theory of Legitimacy. Politics, Prejudice, Social Institutions, and Moral Value". In: *The Psychology of Legiti-*

macy: Emerging Perspectives on Ideology, Justice, and Intergroup Relations. Ed. by John T. Jost and Brenda Major. Cambridge University Press, pp. 77–102.

Cremer, De and Marius van Dijke (2009). "On the Psychology of Justice as a Social Regulation Tool". In: *Netherlands Journal of Psychology* 65.4, pp. 114–117.

Croson, Rachel and Uri Gneezy (2009). "Gender Differences in Preferences". In: *Journal of Economic Literature* 47.2, pp. 448–474.

Croson, Rachel and James Konow (2009). "Social Preferences and Moral Biases". In: *Journal of Economic Behavior & Organization* 69.3, pp. 201–212.

Cushman, Fiery (2008). "Crime and Punishment: Distinguishing the Roles of Causal and Intentional Analyses in Moral Judgment". In: *Cognition* 108.2, pp. 353–380.

Dana, Jason, Roberto A. Weber, and Jason Xi Kuang (2007). "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness". In: *Economic Theory* 33.1, pp. 67–80.

Dankova, Katarina and Maros Servatka (2015). "The House Money Effect and Negative Reciprocity". In: *Journal of Economic Psychology* 48, pp. 60–71.

Darley, John M. and Thomas R. Shultz (1990). "Moral Rules: Their Content and Acquisition". In: *Annual Review of Psychology* 41.1, pp. 525–556.

Demiral, Elif E. and Johanna Mollerstrom (2018). "The Entitlement Effect in the Ultimatum Game – Does It Even Exist?" In: *Journal of Economic Behavior & Organization* 175, pp. 341–352.

DeScioli, Peter, Rebecca Bruening, and Robert Kurzban (2011a). "The Omission Effect in Moral Cognition: Toward a Functional Explanation". In: *Evolution and Human Behavior* 32.3, pp. 204–215.

DeScioli, Peter, John Christner, and Robert Kurzban (2011b). "The Omission Strategy". In: *Psychological Science* 22.4, pp. 442–446.

Dhingra, Nikhil, Zach Gorn, Andrew Kener, and Jason Dana (2012). "The Default Pull: An Experimental Demonstration of Subtle Default Effects on Preferences". In: *Judgment and Decision Making* 7.1, pp. 69–76.

Duch, Raymond M. and Randolph T. Stevenson (2008). *The Economic Vote: How Political and Economic Institutions Condition Election Results*. Cambridge University Press.

Duch, Raymond, Wojtek Przepiorka, and Randolph Stevenson (2014). "Responsibility Attribution for Collective Decision Makers". In: *American Journal of Political Science* 59.2, pp. 372–389.

Duch, Raymond and Randolph Stevenson (Sept. 2013). "Voter Perceptions of Agenda Power and Attribution of Responsibility for Economic Performance". In: *Electoral Studies* 32.3, pp. 512–516.

– (2014). "Pivotal Decision Maker, Agenda Power and Collective Responsibility Attribution". In: *DRAFT*.

Dufwenberg, Martin and Georg Kirchsteiger (2004). "A Theory of Sequential Reciprocity". In: *Games and Economic Behavior* 47.2, pp. 268–298.

Dufwenberg, Martin and Astri Muren (2006). "Gender Composition in Teams". In: *Journal of Economic Behavior & Organization* 61.1, pp. 50–54.

Dworkin, Ronald (1981). "What Is Equality? Part 2: Equality of Resources". In: *Philosophy and Public Affairs*. 4th ser. 10, pp. 283–345.

Eagly, Alice H. (1995). "The Science and Politics of Comparing Women and Men." In: *American Psychologist* 50.3, pp. 145–158.

Eckel, Catherine C. and Philip J. Grossman (1996). "Altruism in Anonymous Dictator Games". In: *Games and Economic Behavior* 16.2, pp. 181–191.

– (1998). "Are Women Less Selfish Than Men?: Evidence From Dictator Experiments". In: *The Economic Journal* 108.448, pp. 726–735.

– (2001). "Chivalry and Solidarity in Ultimatum Games". In: *Economic Inquiry* 39.2, pp. 171–188.

– (2008). "Men, Women and Risk Aversion: Experimental Evidence". In: *Handbook of Experimental Economics Results*. Elsevier, pp. 1061–1073.

Ellemers, Naomi, Henk Wilke, and Ad van Knippenberg (1993). "Effects of the Legitimacy of Low Group or Individual Status on Individual and Collective Status-Enhancement Strategies." In: *Journal of Personality and Social Psychology* 64.5, pp. 766–778.

Elster, Jon (1989). "Social Norms and Economic Theory". In: *Journal of Economic Perspectives* 3.4, pp. 99–117.

Engel, Christoph (2011). "Dictator Games: A Meta Study". In: *Experimental Economics* 14.4, pp. 583–610.

Engel, Christoph and Peter G. Moffatt (2014). "Dhreg, Xtdhreg, and Bootdhreg: Commands to Implement Double-Hurdle Regression". In: *The Stata Journal: Promoting communications on statistics and Stata* 14.4, pp. 778–797.

Engelmann, Dirk and Martin Strobel (2004). "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments". In: *American Economic Review* 94.4, pp. 857–869.

Engl, Florian (2018). "A Theory of Causal Responsibility Attribution". In: *SSRN working paper 2932769*.

Epstein, Richard A. (1973). "A Theory of Strict Liability". In: *The Journal of Legal Studies* 2.1, pp. 151–204.

Erev, Ido, Eyal Ert, and Eldad Yechiam (Dec. 2008). "Loss Aversion, Diminishing Sensitivity, and the Effect of Experience on Repeated Decisions". In: *Journal of Behavioral Decision Making* 21.5, pp. 575–597.

Eriksson, Kimmo and Brent Simpson (2010). "Emotional Reactions to Losing Explain Gender Differences in Entering a Risky Lottery". In: *Judgment and Decision Making* 5.3, pp. 159–163.

Erkut, Hande and Ernesto Reuben (2019). "Preference Measurement and Manipulation in Experimental Economics". In: *Handbook of Research Methods and Applications in Experimental Economics*. Edward Elgar Publishing, pp. 39–56.

Essen, Emma von and Eva Ranehill (2011). "Dominance and Submission: Social Status Biases Economic Sanctions". In: *SSE/EFI Working Paper Series in Economics and Finance* 732. Ed. by The Economic Research Institute (EFI) Stockholm School of Economics.

Falk, Armin (2017). "Status Inequality, Moral Disengagement and Violence". In: *Deutsches Institut für Wirtschaftsforschung (DIW), Berlin* 1676.

Falk, Armin, Ernst Fehr, and Urs Fischbacher (2003). "On the Nature of Fair Behavior". In: *Economic Inquiry* 41.1, pp. 20–26.

– (2005). "Driving Forces Behind Informal Sanctions". In: *Econometrica* 73.6, pp. 2017–2030.

– (2008). "Testing Theories of Fairness—Intentions Matter". In: *Games and Economic Behavior* 62.1, pp. 287–303.

Falk, Armin and Urs Fischbacher (2006). "A Theory of Reciprocity". In: *Games and Economic Behavior* 54.2, pp. 293–315.

Falk, Armin, Thomas Neuber, and Nora Szech (2020). "Diffusion of Being Pivotal and Immoral Outcomes". In: *The Review of Economic Studies* 87.5, pp. 2205–2229.

Falk, Armin and Nora Szech (2013). "Organizations, Diffused Pivotality and Immoral Outcomes". In: *IZA Discussion Papers* No. 7442.

Feess, Eberhard, Florian Kerzenmacher, and Gerd Mühlheußer (2020). "Moral Transgressions by Groups: What Drives Individual Voting Behavior?" In: *CESifo Working Papers*.

Fehr-Duda, Helga, Manuele de Gennaro, and Renate Schubert (2006). "Gender, Financial Risk, and Probability Weights". In: *Theory and Decision* 60.2-3, pp. 283–313.

Fehr, Ernst and Urs Fischbacher (2004). “Third-Party Punishment and Social Norms”. In: *Evolution and Human Behavior* 25.2, pp. 63–87.

Fehr, Ernst and Simon Gächter (2000a). “Cooperation and Punishment in Public Goods Experiments”. In: *American Economic Review* 90.4, pp. 980–994.

– (2000b). “Fairness and Retaliation: The Economics of Reciprocity”. In: *Journal of Economic Perspectives* 14.3, pp. 159–182.

– (2002). “Altruistic Punishment in Humans”. In: *Nature* 415.6868, pp. 137–140.

Fehr, Ernst and Herbert Gintis (2007). “Human Motivation and Social Cooperation: Experimental and Analytical Foundations”. In: *Annual Review of Sociology* 33.1, pp. 43–64.

Fehr, Ernst and Klaus M. Schmidt (1999). “A Theory of Fairness, Competition, and Cooperation”. In: *The Quarterly Journal of Economics* 114.3, pp. 817–868.

Feltovich, Nick (2019). “Is Earned Bargaining Power more Fully Exploited?” In: *Journal of Economic Behavior & Organization* 167, pp. 152–180.

Fernandez, Raquel and Dani Rodrik (1991). “Resistance to Reform: Status Quo Bias in the Presence of Individual-Specific Uncertainty”. In: *American Economic Review* 81.5, pp. 1146–1155.

Fershtman, Chaim (1996). “On the Value of Incumbency Managerial Reference Points and Loss Aversion”. In: *Journal of Economic Psychology* 17.2, pp. 245–257.

Fershtman, Chaim and Uri Gneezy (2001). “Strategic Delegation: An Experiment”. In: *The RAND Journal of Economics* 32.2, pp. 352–368.

Fershtman, Chaim, Uri Gneezy, and John A. List (2012). “Equity Aversion: Social Norms and the Desire to be Ahead”. In: *American Economic Journal: Microeconomics* 4.4, pp. 131–144.

Fiorina, Morris P. (1982). “Legislative Choice of Regulatory Forms: Legal Process or Administrative Process?” In: *Public Choice* 39, pp. 33–66.

Fischbacher, Urs and Franziska Föllmi-Heusi (2013). “Lies in Disguise - An Experimental Study on Cheating”. In: *Journal of the European Economic Association* 11.3, pp. 525–547.

Fischbacher, Urs and Simeon Schudy (2014). “Reciprocity and Resistance to Comprehensive Reform”. In: *Public Choice* 160.3-4, pp. 411–428.

– (2020). “Agenda Control and Reciprocity in Sequential Voting Decisions”. In: *Economic Inquiry* 58.4, pp. 1813–1829.

Fleiß, Jürgen (2015). “Merit Norms in the Ultimatum Game: An Experimental Study of the Effect of Merit on Individual Behavior and Aggregate Outcomes”. In: *Central European Journal of Operations Research* 23.2, pp. 389–406.

Folger, Robert (1987). "Reformulating the Preconditions of Resentment: A Referent Cognitions Model". In: *Social Comparison, Social Justice, and Relative Deprivation: Theoretical, Empirical, and Policy Perspectives*. Ed. by John C. Masters and William P. Smith. Lawrence Erlbaum Associates, Inc., pp. 183–215.

Fong, Christina (2001). "Social Preferences, Self-Interest, and the Demand for Redistribution". In: *Journal of Public Economics* 82.2, pp. 225–246.

Franzese, Robert J. (2002). *Macroeconomic Policies of Developed Democracies*. Cambridge University Press.

French, John R. P. and Bertram Raven (1959). "The Bases of Social Power". In: *Studies in Social Power*. Ed. by Dorwin Cartwright. The University of Michigan, Ann Arbor, pp. 150–167.

Frey, Bruno S. (1997). *Not Just for the money. An Economic Theory of Personal Motivation*. Edward Elgar Publishing Limited.

Frey, Bruno S., Matthias Benz, and Alois Stutzer (2004). "Introducing Procedural Utility: Not Only What, but Also How Matters". In: *Journal of Institutional and Theoretical Economics (JITE)* 160.3, pp. 377–401.

Frey, Bruno S. and Lorenz Götte (1999). "Does Pay Motivate Volunteers". In: *Unpublished Manuscript. Institute for Empirical Economic Research*.

Fritz, Annemarie, Walter Hussy, and David Tobinski (2018). *Pädagogische Psychologie*. utb GmbH.

Frohlich, Norman and Joe Oppenheimer (1994). "Preferences for Income Distribution and Distributive Justice: A Window on the Problems of Using Experimental Data in Economics and Ethics". In: *Eastern Economic Journal* 20.2, pp. 147–155.

Frohlich, Norman, Joe Oppenheimer, and Anja Kurki (2004). "Modeling Other-Regarding Preferences and an Experimental Test". In: *Public Choice* 119.1/2, pp. 91–117.

Gächter, Simon, Leonie Gerhards, and Daniele Nosenzo (2017). "The Importance of Peers for Compliance with Norms of Fair Sharing". In: *European Economic Review* 97, pp. 72–86.

Gächter, Simon, Eric J. Johnson, and Andreas Herrmann (2021). "Individual-Level Loss Aversion in Riskless and Risky Choices". In: *Theory and Decision* 92.3-4, pp. 599–624.

Gächter, Simon and Arno Riedl (2005). "Moral Property Rights in Bargaining with Infeasible Claims". In: *Management Science* 51.2, pp. 249–263.

Gailey, Jeannine A. and R. Frank Falk (2008). "Attribution of Responsibility as a Multidimensional Concept". In: *Sociological Spectrum* 28.6, pp. 659–680.

Gärtner, Manja (2018). "The Prosociality of Intuitive Decisions Depends on the Status Quo". In: *Journal of Behavioral and Experimental Economics* 74, pp. 127–138.

Gärtner, Manja and Anna Sandberg (2017). “Is there an Omission Effect in Prosocial Behavior? A Laboratory Experiment on Passive vs. Active Generosity”. In: *PloS one* 12.3. Ed. by Eldad Yechiam, e0172496.

Gelman, Andrew, Jonathan N. Katz, and Joseph Bafumi (2004). “Standard Voting Power Indexes Do Not Work: An Empirical Analysis”. In: *British Journal of Political Science* 34.4, pp. 657–674.

Gelman, Andrew, Jonathan N. Katz, and Francis Tuerlinckx (2002). “The Mathematics and Statistics of Voting Power”. In: *Statistical Science* 17.4, pp. 420–435.

Gerstenberg, Tobias, Tomer D. Ullman, Jonas Nagel, Max Kleiman-Weiner, David A. Lagnado, and Joshua B. Tenenbaum (2018). “Lucky or Clever? From Expectations to Responsibility Judgments”. In: *Cognition* 177, pp. 122–141.

Gill, David and Victoria Prowse (2019). “Measuring Costly Effort Using the Slider Task”. In: *Journal of Behavioral and Experimental Finance* 21, pp. 1–9.

Gill, David and Rebecca Stone (2015). “Desert and Inequity Aversion in Teams”. In: *Journal of Public Economics* 123, pp. 42–54.

Gimbel, Ronald W., Martin A. Strosberg, Susan E. Lehrman, Eugenijus Gefenas, and Frank Taft (2003). “Presumed Consent and other Predictors of Cadaveric Organ Donation in Europe”. In: *Progress in Transplantation* 13.1, pp. 17–23.

Gneezy, Uri and Aldo Rustichini (2000). “Pay Enough or Don’t Pay at All”. In: *Quarterly Journal of Economics* 115.3, pp. 791–810.

Gold, Andreas and Elmar Souvignier (2005). “Prognose der Studierfähigkeit”. In: *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 37.4, pp. 214–222.

Gomez, Brad T. and J. Matthew Wilson (2003). “Causal Attribution and Economic Voting in American Congressional Elections”. In: *Political Research Quarterly* 56.3, pp. 271–282.

Granger, Clive W. J. (1969). “Investigating Causal Relations by Econometric Models and Cross-spectral Methods”. In: *Econometrica* 37.3, p. 424.

– (1988). “Some Recent Development in a Concept of Causality”. In: *Journal of Econometrics* 39.1-2, pp. 199–211.

Grossman, Zachary (2014). “Strategic Ignorance and the Robustness of Social Preferences”. In: *Management Science* 60.11, pp. 2659–2665.

Gurdal, Mehmet Y., Joshua B. Miller, and Aldo Rustichini (2013). “Why Blame?” In: *Journal of Political Economy* 121.6, pp. 1205–1247.

Güth, Werner, Steffen Huck, and Wieland Müller (2001). “The Relevance of Equal Splits in Ultimatum Games”. In: *Games and Economic Behavior* 37.1, pp. 161–169.

Hamburg, Universität (2022). *NC-Werte (ohne Lehrämter)*. last retrieved 21.12.2022. URL: https://www.uni-hamburg.de/campuscenter/studienorganisation/formulare-informationsmerkblaetter/nc-wartesemester.pdf.

Hamman, John R., George Loewenstein, and Roberto A. Weber (2010). "Self-Interest through Delegation: An Additional Rationale for the Principal-Agent Relationship". In: *American Economic Review* 100.4, pp. 1826–1846.

Hart, H.L.A. and Tony Honoré (1985). *Causation in the Law*. Oxford University Press Inc., New York.

Hausman, Daniel M. (1986). "Causation and Experimentation". In: *American Philosophical Quarterly* 23.2, pp. 143–154.

Hayashi, Andrew T. (2013). "Occasionally Libertarian: Experimental Evidence of Self-Serving Omission Bias". In: *Journal of Law, Economics, and Organization* 29.3, pp. 711–733.

Hays, Nicholas A. and Steven L. Blader (2017). "To Give or not to Give? Interactive Effects of Status and Legitimacy on Generosity." In: *Journal of Personality and Social Psychology* 112.1, pp. 17–38.

Healy, Andrew and Neil Malhotra (2013). "Retrospective Voting Reconsidered". In: *Annual Review of Political Science* 16.1, pp. 285–306.

Heap, Shaun P. Hargreaves and Daniel John Zizzo (2009). "The Value of Groups". In: *American Economic Review* 99.1, pp. 295–323.

Heffetz, Ori (2011). "A Test of Conspicuous Consumption: Visibility and Income Elasticities". In: *Review of Economics and Statistics* 93.4, pp. 1101–1117.

Heffetz, Ori and Robert H. Frank (2011). "Preferences for Status". In: *Handbook of Social Economics*. Elsevier, pp. 69–91.

Hegtvedt, Karen A., Jody Clay-Warner, and Cathryn Johnson (2003). "The Social Context of Responses to Injustice: Considering the Indirect and Direct Effects of Group-Level Factors". In: *Social Justice Research* 16.4, pp. 343–366.

Heider, Fritz (1958). *The Psychology of Interpersonal Relations*. John Wiley & Sons Inc.

Heine, Christoph, Heiko Quast, and Mareike Beuße (2010). "Studienberechtigte 2008 ein halbes Jahr nach Studienabschluss - Übergang in Studium, Beruf und Ausbildung". In: *HIS: Forum Hochschule* 3.

Henrich, Joseph and Francisco J. Gil-White (2001). "The Evolution of Prestige: Freely Conferred Deference as a Mechanism for Enhancing the Benefits of Cultural Transmission". In: *Evolution and Human Behavior* 22.3, pp. 165–196.

Henrich, Joseph, Richard McElreath, Abigail Barr, Jean Ensminger, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich,

Carolyn Lesorogol, Frank Marlowe, David Tracer, and John Ziker (2006). "Costly Punishment Across Human Societies". In: *Science* 312.5781, pp. 1767–1770.

Herrmann, Benedikt, Christian Thoöni, and Simon Gaächter (2008). "Antisocial Punishment Across Societies". In: *Science* 319.5868, pp. 1362–1367.

Hett, Florian, Mario Mechtel, and Markus Kröll (2020). "The Structure and Behavioral Effects of Revealed Social Identity Preferences". In: *The Economic Journal* 130.632, pp. 2569–2595.

Hewstone, Miles, Mark Rubin, and Hazel Willis (2002). "Intergroup Bias". In: *Annual Review of Psychology* 53.1, pp. 575–604.

Hinsch, Wilfried (2008). "Legitimacy and Justice". In: *Political Legitimization without Morality?* Springer Netherlands, pp. 39–52.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith (1994). "Preferences, Property Rights, and Anonymity in Bargaining Games". In: *Games and Economic Behavior* 7, pp. 346–380.

Hoffman, Elizabeth and Matthew L. Spitzer (1985). "Entitlements, Rights, and Fairness: An Experimental Examination of Subjects' Concepts of Distributive Justice". In: *The Journal of Legal Studies* 14.2, pp. 259–297.

Holt, Charles A. and Susan K. Laury (2002). "Risk Aversion and Incentive Effects". In: *American Economic Review* 92.5, pp. 1644–1655.

Homans, George C. (1958). "Social Behavior as Exchange". In: *American Journal of Sociology* 63.6, pp. 597–606.

Hong, Kessely and Iris Bohnet (2007). "Status and Distrust: The Relevance of Inequality and Betrayal Aversion". In: *Journal of Economic Psychology* 28.2, pp. 197–213.

Hough, Mike, Jonathan Jackson, and Ben Bradford (2013). "Legitimacy, Trust and Compliance: An Empirical Test of Procedural Justice Theory using the European Social Survey". In: *Legitimacy and Criminal Justice: An International Exploration*. Ed. by Justice Tankebe. Oxford University Press, pp. 326–352.

Huberman, Bernardo A., Christoph H. Loch, and Ayse Önçüler (2004). "Status As a Valued Resource". In: *Social Psychology Quarterly* 67.1, pp. 103–114.

Hülle, Sebastian, Stefan Liebig, and Meike Janina May (2018). "Measuring Attitudes Toward Distributive Justice: The Basic Social Justice Orientations Scale". In: *Social Indicators Research* 136.2, pp. 663–692.

Hume, David (1948). *An Inquiry Concerning Human Understanding*. Bobbs Merrill. Originally published 1748.

Ingenkamp, Karl-Heinz (1997). *Lehrbuch der Pädagogischen Diagnostik*. Weihnheim: Beltz.

Isaac, R. Mark, James M. Walker, and Susan H. Thomas (1984). "Divergent Evidence on Free Riding: An Experimental Examination of Possible Explanations". In: *Public Choice* 43.2, pp. 113–149.

Jachimowicz, Jon M., Shannon Duncan, Elke U. Weber, and Eric J. Johnson (2019). "When and Why Defaults Influence Decisions: A Meta-Analysis of Default Effects". In: *Behavioural Public Policy* 3.02, pp. 159–186.

Jasso, Guillermina, Kjell Y. Törnblom, and Clara Sabbagh (2016). "Distributive Justice". In: *Handbook of Social Justice Theory and Research*. Springer New York, pp. 201–218.

Jerit, Jennifer, Jason Barabas, and Toby Bolsen (2006). "Citizens, Knowledge, and the Information Environment". In: *American Journal of Political Science* 50.2, pp. 266–282.

Johnson, Eric J. and Daniel Goldstein (2003). "Do Defaults Save Lives?" In: *Science* 302.5649, pp. 1338–1339.

Johnson, Eric J., John Hershey, Jacqueline Meszaros, and Howard Kunreuther (1993). "Framing, Probability distortions, and Insurance Decisions". In: *Journal of Risk and Uncertainty* 7.1, pp. 35–51.

Jost, John T. and Brenda Major (2001). "Emerging Perspectives on the Psychology of Legitimacy". In: *The Psychology of Legitimacy: Emerging Perspectives on Ideology, Justice, and Intergroup Relations*. Ed. by John T. Jost and Brenda Major. Cambridge University Press, pp. 3–30.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler (1991). "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias". In: *Journal of Economic Perspectives* 5.1, pp. 193–206.

Kahneman, Daniel and Dale T. Miller (1986). "Norm Theory: Comparing Reality to Its Alternatives." In: *Psychological Review* 93.2, pp. 136–153.

Kahneman, Daniel and Amos Tversky (1979). "Prospect Theory: An Analysis of Decision under Risk". In: *Econometrica* 47.2, pp. 263–292.

– (1982). "The Psychology of Preferences". In: *Scientific American* 246.1, pp. 160–173.

Kalkhoff, Will and Christopher Barnum (2000). "The Effects of Status-Organizing and Social Identity Processes on Patterns of Social Influence". In: *Social Psychology Quarterly* 63.2, pp. 95–115.

Katzenbach, Jon R. and Douglas K. Smith (1994). *The Wisdom of Teams: Creating the High-Performance Organization*. Harvard Business Review Press. Boston, Massachusetts.

Kayser, Mark A. and Michael Peress (2012). "Benchmarking across Borders: Electoral Accountability and the Necessity of Comparison". In: *American Political Science Review* 106.3, pp. 661–684.

Kelley, Jonathan and M. D. R. Evans (1993). "The Legitimation of Inequality: Occupational Earnings in Nine Nations". In: *American Journal of Sociology* 99.1, pp. 75–125.

Kelman, Herbert C. (2001). "Reflections on Social and Psychological Processes of Legitimization and Delegitimization". In: *The Psychology of Legitimacy: Emerging Perspectives on Ideology, Justice, and Intergroup Relations*. Ed. by John T. Jost and Brenda Major. Cambridge University Press, pp. 54–73.

Kinder, Donald R. and D. Roderick Kiewiet (1981). "Sociotropic Politics: The American Case". In: *British Journal of Political Science* 11.2, pp. 129–161.

King, Gary (1991). "Constituency Service and Incumbency Advantage". In: *British Journal of Political Science* 21.1, pp. 119–128.

Knight, Carl (2009). *Luck Egalitarianism. Equality, Responsibility, and Justice*. Edinburgh University Press, Edinburgh.

– (2013). "Luck Egalitarianism". In: *Philosophy Compass* 8.10, pp. 924–934.

Konferenz, Kultusminister (2021). *Abiturnoten im Ländervergleich*. last retrieved 21.12.2022. URL: `https://www.kmk.org/dokumentation-statistik/statistik/schulstatistik/abiturnoten.html`.

Konow, James (2000). "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions". In: *American Economic Review* 90.4, pp. 1072–1091.

Konow, James and Lars Schwettmann (2016). "The Economics of Justice". In: *Handbook of Social Justice Theory and Research*. Springer New York, pp. 83–106.

Kramer, Gerald H. (1983). "The Ecological Fallacy Revisited: Aggregate- versus Individual-level Findings on Economics and Elections, and Sociotropic Voting". In: *American Political Science Review* 77.1, pp. 92–111.

Kraus, Michael W., Paul K. Piff, and Dacher Keltner (2009). "Social Class, Sense of Control, and Social Explanation." In: *Journal of Personality and Social Psychology* 97.6, pp. 992–1004.

Krawczyk, Michał (2010). "A Glimpse Through the Veil of Ignorance: Equality of Opportunity and Support for redistribution". In: *Journal of Public Economics* 94.1-2, pp. 131–141.

Krupka, Erin L. and Roberto A. Weber (2013). "Identifying Social Norms Using Coordination Games: Why does Dictator Game Sharing Vary?" In: *Journal of the European Economic Association* 11.3, pp. 495–524.

Ku, Hyejin and Timothy C. Salmon (2013). "Procedural Fairness and the Tolerance for Income Inequality". In: *European Economic Review* 64, pp. 111–128.

Kugler, Tamar, Edgar E. Kausel, and Martin G. Kocher (2012). "Are Groups More Rational than Individuals? A Review of Interactive Decision Making in Groups". In: *WIREs Cognitive Science* 3.4, pp. 471–482.

Laibson, David (1997). "Golden Eggs and Hyperbolic Discounting". In: *The Quarterly Journal of Economics* 112.2, pp. 443–478.

Landman, Janet (1987). "Regret and Elation Following Action and Inaction". In: *Personality and Social Psychology Bulletin* 13.4, pp. 524–536.

Langenhoff, Antonia F., Alex Wiegmann, Joseph Y. Halpern, Joshua B. Tenenbaum, and Tobias Gerstenberg (2021). "Predicting Responsibility Judgments from Dispositional Inferences and Causal Attributions". In: *Cognitive Psychology* 129, p. 101412.

Leibbrandt, Andreas and Raúl López-Pérez (2012). "An Exploration of Third and Second Party Punishment in Ten Simple Games". In: *Journal of Economic Behavior & Organization* 84.3, pp. 753–766.

Lerner, Melvin J. (1987). "Integrating Societal and Psychological Rules of Entitlement: The Basic Task of Each Social Actor and Fundamental Problem for the Social Sciences". In: *Social Justice Research* 1.1, pp. 107–125.

Levin, Shana, Christopher M. Federico, Jim Sidanius, and Joshua L. Rabinowitz (2002). "Social Dominance Orientation and Intergroup Bias: The Legitimation of Favoritism for High-Status Groups". In: *Personality and Social Psychology Bulletin* 28.2, pp. 144–157.

Levine, David K. (1998). "Modeling Altruism and Spitefulness in Experiments". In: *Review of Economic Dynamics* 1.3, pp. 593–622.

Levitt, Steven D. and John A. List (2009). "Field Experiments in Economics: The Past, the Present, and the Future". In: *European Economic Review* 53.1, pp. 1–18.

Lewis-Beck, Michael S. (1988). *Economics and Elections: The Major Western Democracies*. The University of Michigan Press.

List, John A. (2007). "On the Interpretation of Giving in Dictator Games". In: *Journal of Political Economy* 115.3, pp. 482–493.

Loomes, Graham and Robert Sugden (1982). "Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty". In: *The Economic Journal* 92.368, p. 805.

Luhan, Wolfgang J., Martin G. Kocher, and Matthias Sutter (2009). "Group Polarization in the Team Dictator Game Reconsidered". In: *Experimental Economics* 12.1, pp. 26–41.

Mackie, J. L. (1974). *The Cement of the Universe: A Study of Causation*. Oxford University Press.

Madrian, Brigitte C. and Dennis F. Shea (2001). "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior". In: *The Quarterly Journal of Economics* 116.4, pp. 1149–1187.

Malle, Bertram F., Steve Guglielmo, and Andrew E. Monroe (2014). "A Theory of Blame". In: *Psychological Inquiry* 25.2, pp. 147–186.

Mandel, David R. and Darrin R. Lehman (1996). "Counterfactual Thinking and Ascriptions of Cause and Preventability." In: *Journal of Personality and Social Psychology* 71.3, pp. 450–463.

Maniadis, Zacharias, Fabio Tufano, and John A. List (2017). "To Replicate or Not to Replicate? Exploring Reproducibility in Economics through the Lens of a Model and a Pilot Study". In: *The Economic Journal* 127.605, pp. 209–235.

Markus, Gregory B. (1988). "The Impact of Personal and National Economic Conditions on the Presidential Vote: A Pooled Cross-Sectional Analysis". In: *American Journal of Political Science* 32.1, pp. 137–154.

McKenzie, Craig R. M., Michael J. Liersch, and Stacey R. Finkelstein (2006). "Recommendations Implicit in Policy Defaults". In: *Psychological Science* 17.5, pp. 414–420.

Mertins, Vanessa, Henrik Egbert, and Tanja Könen (2013). "The Effects of Individual Judgments about Selection Procedures: Results from a Power-to-Resist Game". In: *The Journal of Socio-Economics* 42, pp. 112–120.

Mikula, Gerold (2003). "Testing an Attribution-of-Blame Model of Judgments of Injustice". In: *European Journal of Social Psychology* 33.6, pp. 793–811.

Miller, Dale (2001). "Disrespect and the Experience of Injustice". In: *Annual Review of Psychology* 52, pp. 527–553.

Miller, David (1999). *Principles of Social Justice*. Harvard University Press, Cambridge.

Molenmaker, Welmer E., Erik W. de Kwaadsteniet, and Eric van Dijk (2016). "The Impact of Personal Responsibility on the (Un)willingness to Punish Non-Cooperation and Reward Cooperation". In: *Organizational Behavior and Human Decision Processes* 134, pp. 1–15.

Mollerstrom, Johanna, Bjørn-Atle Reme, and Erik Ø. Sørensen (2015). "Luck, Choice and Responsibility — An Experimental Study of Fairness Views". In: *Journal of Public Economics* 131, pp. 33–40.

Mullen, Brian, Rupert Brown, and Colleen Smith (1992). "Ingroup Bias as a Function of Salience, Relevance, and Status: An Integration". In: *European Journal of Social Psychology* 22.2, pp. 103–122.

Nelson, Julie A. (2014). "Are Women really more Risk-Averse than Men? A Re-analysis of the Literature using Expended Methods". In: *Journal of Economic Surveys* 29.3, pp. 566–585.

O'Donoghue, Ted and Matthew Rabin (1999). "Doing It Now or Later". In: *American Economic Review* 89.1, pp. 103–124.

Oexl, Regine and Zachary J. Grossman (2012). "Shifting the Blame to a Powerless Intermediary". In: *Experimental Economics* 16.3, pp. 306–312.

Offerman, Theo (2002). "Hurting Hurts more than Helping Helps". In: *European Economic Review* 46.8, pp. 1423–1437.

Olson, James M. and Carolyn L. Hafer (2001). "Tolerance of Personal Deprivation". In: *The Psychology of Legitimacy: Emerging Perspectives on Ideology, Justice, and Intergroup Relations*. Ed. by John T. Jost and Brenda Major. Cambridge University Press, pp. 157–175.

Orcutt, Guy H. (1952). "Actions, Consequences, and Causal Relations". In: *The Review of Economics and Statistics* 34.4, p. 305.

Ostrom, Elinor (2000). "Collective Action and the Evolution of Social Norms". In: *Journal of Economic Perspectives* 14.3, pp. 137–158.

Oxoby, Robert J. and Kendra N. McLeish (2004). "Sequential Decision and Strategy Vector Methods in Ultimatum Bargaining: Evidence on the Strength of Other-Regarding Behavior". In: *Economics Letters* 84.3, pp. 399–405.

Paetzel, Fabian and Rupert Sausgruber (2018). "Cognitive Ability and In-Group Bias: An Experimental Study". In: *Journal of Public Economics* 167, pp. 280–292.

Panchanathan, Karthik, Willem E. Frankenhuis, and Joan B. Silk (2013). "The Bystander Effect in an N-Person Dictator Game". In: *Organizational Behavior and Human Decision Processes* 120.2, pp. 285–297.

Pareto, Vilfredo (1906). *Manual of Political Economy*. New York: Augustus M. Kelley.

Penrose, Lionel S. (1946). "The Elementary Statistics of Majority Voting". In: *Journal of the Royal Statistical Society* 109.1, pp. 53–57.

Piaget, Jean (1932). *The Moral Judgment of the Child*. New York: Harcourt, Brace.

Pichert, Daniel and Konstantinos V. Katsikopoulos (2008). "Green Defaults: Information Presentation and Pro-Environmental Behaviour". In: *Journal of Environmental Psychology* 28.1, pp. 63–73.

Piff, Paul K., Michael W. Kraus, Stéphane Côté, Bonnie Hayden Cheng, and Dacher Keltner (2010). "Having Less, Giving More: The Influence of Social Class on Prosocial Behavior." In: *Journal of Personality and Social Psychology* 99.5, pp. 771–784.

Powell, G. Bingham and Guy D. Whitten (1993). "A Cross-National Analysis of Economic Voting: Taking Account of the Political Context". In: *American Journal of Political Science* 37.2, pp. 391–414.

Price, Vincent and John Zaller (1993). "Who Gets the News? Alternative Measures of News Reception and Their Implications for Research". In: *Public Opinion Quarterly* 57.2, p. 133.

Putterman, Louis, Jean-Robert Tyran, and Kenju Kamei (2011). "Public Goods and Voting on Formal Sanction Schemes". In: *Journal of Public Economics* 95.9-10, pp. 1213–1222.

Quattrone, George A. and Amos Tversky (1988). "Contrasting Rational and Psychological Analyses of Political Choice". In: *American Political Science Review* 82.3, pp. 719–736.

Rabin, Matthew (1993). "Incorporating Fairness into Game Theory and Economics". In: *American Economic Review* 83.5, pp. 1281–1302.

Raftery, Adrian E. (1995). "Bayesian Model Selection in Social Research". In: *Sociological Methodology* 25, p. 111.

Ramalingam, Abhijit, Antonio J. Morales, and James M. Walker (2019). "Peer Punishment of Acts of Omission Versus Acts of Commission in Give and Take Social Dilemmas". In: *Journal of Economic Behavior & Organization* 164, pp. 133–147.

Rauhut, Heiko and Fabian Winter (2010). "A Sociological Perspective on Measuring Social Norms by Means of Strategy Method Experiments". In: *Social Science Research* 39.6, pp. 1181–1194.

Rawls, John (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Ridgeway, Cecilia L. (2001). "The Emergence of Status Beliefs. From Structural Inequality to Legitimizing Ideology". In: *The Psychology of Legitimacy: Emerging Perspectives on Ideology, Justice, and Intergroup Relations*. Ed. by John T. Jost and Brenda Major. Cambridge University Press, pp. 257–277.

Ridgeway, Cecilia L., Elizabeth Heger Boyle, Kathy J. Kuipers, and Dawn T. Robinson (1998). "How Do Status Beliefs Develop? The Role of Resources and Interactional Experience". In: *American Sociological Review* 63.3, pp. 331–350.

Ritov, Ilana and Jonathan Baron (1992). "Staus-Quo and Omission Bias". In: *Journal of Risk and Uncertainty* 5, pp. 49–61.

Roth, Alvin E. (1995). "Bargaining Experiments". In: *The Handbook of Experimental Economics*. Ed. by John H. Kagel and Alvin E. Roth. Princeton University Press, pp. 253–348.

Rothenhäusler, Dominik, Nikolaus Schweizer, and Nora Szech (2013). "Institutions, Shared Guilt and Moral Transgression". In: *CESifo Working Paper Series* 5525.

Samuelson, William and Richard Zeckhauser (1988). "Status Quo Bias in Decision Making". In: *Journal of Risk and Uncertainty* 1.1, pp. 7–59.

Sandroni, Alvaro, Sandra Ludwig, and Philipp Kircher (2013). "On the Difference between Social and Private Goods". In: *The B.E. Journal of Theoretical Economics* 13.1, pp. 151–177.

Schlenker, Barry R., Thomas W. Britt, John Pennington, Rodolfo Murphy, and Kevin Doherty (1994). "The Triangle Model of Responsibility." In: *Psychological Review* 101.4, pp. 632–652.

Schlicht, Ekkehart (Mar. 1998). *On Custom in the Economy*. Oxford University Press.

Schmidt, Ulrich and Stefan Traub (2002). "An Experimental Test of Loss Aversion". In: *Journal of Risk and Uncertainty* 25.3, pp. 233–249.

Schokkaert, Erik and Kurt Devooght (2003). "Responsibility-Sensitive Fair Compensation in Different Cultures". In: *Social Choice and Welfare* 21.2, pp. 207–242.

Schweitzer, Maurice (1994). "Disentangling Status Quo and Omission Effects: An Experimental Analysis". In: *Organizational Behavior and Human Decision Processes* 58.3, pp. 457–476.

Sebald, Alexander (2010). "Attribution and Reciprocity". In: *Games and Economic Behavior* 68.1, pp. 339–352.

Segal, Carmit (2012). "Working When No One Is Watching: Motivation, Test Scores, and Economic Success". In: *Management Science* 58.8, pp. 1438–1457.

Selten, Reinhard (1967). "Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes". In: *Beiträge zur experimentellen Wirtschaftsforschung*. Ed. by Heinz Sauermann. J.C.B Mohr (Paul Siebeck).

Selten, Reinhard and Axel Ockenfels (1998). "An Experimental Solidarity Game". In: *Journal of Economic Behavior & Organization* 34.4, pp. 517–539.

Serra-Garcia, Marta and Nora Szech (2022). "Incentives and Defaults Can Increase COVID-19 Vaccine Intentions and Test Demand". In: *Management Science*.

Shapley, Lloyd S. and Martin Shubik (1954). "A Method for Evaluating the Distribution of Power in a Committee System". In: *American Political Science Review* 48.3, pp. 787–792.

Shaver, Kelly G. (1985). *The Attribution of Blame: Causality, Responsibility, and Blameworthiness*. New York: Springer-Verlag.

Shayo, Moses (2009). "A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution". In: *American Political Science Review* 103.2, pp. 147–174.

Shih, Margaret, Todd L. Pittinsky, and Nalini Ambady (1999). "Stereotype Susceptibility: Identity Salience and Shifts in Quantitative Performance". In: *Psychological Science* 10.1, pp. 80–83.

Shultz, Thomas R., Michael Schleifer, and Ian Altman (1981). "Judgments of Causation, Responsibility, and Punishment in Cases of Harm-Doing." In: *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement* 13.3, pp. 238–253.

Siebel, Mark and Thomas Schramme (2020). "Need-Based Justice from the Perspective of Philosophy". In: *Need-Based Distributive Justice*. Ed. by Stefan Traub and Bernhard Kittel. Springer International Publishing, pp. 21–58.

Simpson, Brent and Henry A. Walker (2002). "Status Characteristics and Performance Expectations: A Reformulation". In: *Sociological Theory* 20.1, pp. 24–40.

Singer, Tania, Ben Seymour, John P. O'Doherty, Klaas E. Stephan, Raymond J. Dolan, and Chris D. Frith (2006). "Empathic Neural Responses are Modulated by the Perceived Fairness of Others". In: *Nature* 439.7075, pp. 466–469.

Sobel, Joel (2005). "Interdependent Preferences and Reciprocity". In: *Journal of Economic Literature* 43.2, pp. 392–436.

Solnick, Sara J. (2001). "Gender Differences in the Ultimatum Game". In: *Economic Inquiry* 39.2, pp. 189–200.

Spellman, Barbara A. (1997). "Crediting Causality". In: *Journal of Experimental Psychology: General* 126.4, pp. 323–348.

Spranca, Mark, Elisa Minsk, and Jonathan Baron (1991). "Omission and Commission in Judgment and Choice". In: *Journal of Experimental Social Psychology* 27.1, pp. 76–105.

Sunstein, Cass R. and Lucia A. Reisch (2021). "Climate-Friendly Default Rules". In: *Sustainable Consumption and Production, Volume I*. Springer International Publishing, pp. 141–164.

Süß, Heinz-Martin (2001). "Prädiktive Validität der Intelligenz im schulischen und außerschulischen Bereich". In: *Perspektiven der Intelligenzforschung*. Ed. by E. Stern. Lengerich: Pabst, pp. 109–136.

Tajfel, Henri, M. G. Billig, R. P. Bundy, and Claude Flament (1971). "Social Categorization and Intergroup Behaviour". In: *European Journal of Social Psychology* 1.2, pp. 149–178.

Tajfel, Henri and John C. Turner (1979). "An Integrative Theory of Intergroup Conflict". In: *The social psychology of intergroup relations*. Ed. by William G. Austin, pp. 33–47. Monterey, California: Brooks/Cole.

Thaler, Richard H. and Cass R. Sunstein (2003). "Libertarian Paternalism". In: *American Economic Review* 93.2, pp. 175–179.

Thibaut, John W. and Laurens Walker (1975). *Procedural Justice: A Psychological Analysis*. Hillsdale, NJ: Erlbaum.

Tost, Leigh Plunkett (2011). "An Integrative Model of Legitimacy Judgments". In: *Academy of Management Review* 36.4, pp. 686–710.

Traub, Stefan (2020). "Perspectives for a Theory of Need-Based Distributive Justice". In: *Need-Based Distributive Justice*. Springer International Publishing, pp. 1–20.

Traub, Stefan and Bernhard Kittel, eds. (2020). *Need-Based Distributive Justice. An Interdisciplinary Perspective*. Springer International Publishing.

Trueblood, Jennifer S. (2015). "Reference Point Effects in Riskless Choice Without Loss Aversion." In: *Decision* 2.1, pp. 13–26.

Turner and Brown (1978). "Status, Position, and Legitimacy in Intergroup Behaviour: The Effects of Different Status Relationships on In-Group Bias and Group Creativity". In: *Differentiation between social groups: Studies in the social psychology of intergroup relations*. London: Academic Press Inc., pp. 201–234.

Tutic, Andreas and Ulf Liebe (2009). "A Theory of Status-Mediated Inequity Aversion". In: *The Journal of Mathematical Sociology* 33.3, pp. 157–195.

Tyler, Tom R. (2001). "A Psychological Perspective on the Legitimacy of Institutions and Authorities". In: *The Psychology of Legitimacy: Emerging Perspectives on Ideology, Justice, and Intergroup Relations*. Ed. by John T. Jost and Brenda Major. Cambridge University Press, pp. 416–436.

– (ja 2006). "Psychological Perspectives on Legitimacy and Legitimation". In: *Annual Review of Psychology* 57.1, pp. 375–400.

Vaal, Johanna H. Kordes-de (1996). "Intention and the Omission Bias: Omissions Perceived as Nondecisions". In: *Acta Psychologica* 93.1-3, pp. 161–172.

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

Veblen, Thorstein (1899). "The Theory of the Leisure Class". In: *Journal of Political Economy* 8.1, pp. 106–117.

Vermunt, Riel and Herman Steensma (2016). "Procedural Justice". In: *Handbook of social justice theory and research*. Ed. by Clara Sabbagh and Manfred Schmitt. Springer, New York, pp. 219–236.

Voyer, Daniel and Susan D. Voyer (2014). "Gender Differences in Scholastic Achievement: A Meta-Analysis". In: *Psychological Bulletin* 140.4, pp. 1174–1204.

Weber, Ulrike, Amélie Mummendey, and Sven Waldzus (2002). "Perceived Legitimacy of Intergroup Status Differences: its Prediction by Relative Ingroup Prototypicality". In: *European Journal of Social Psychology* 32.4, pp. 449–470.

Weiner, Bernard (1993). "On Sin Versus Sickness: A Theory of Perceived Responsibility and Social Motivation." In: *American Psychologist* 48.9, pp. 957–965.

Weiss, Yoram and Chaim Fershtman (1998). "Social Status and Economic Performance:" in: *European Economic Review* 42.3-5, pp. 801–820.

Wells, Gary L. and Igor Gavanski (1989). "Mental Simulation of Causality". In: *Journal of Personality and Social Psychology* 56.2, pp. 161–169.

Wentzel, Kathryn R. (1991). "Relations between Social Competence and Academic Achievement in Early Adolescence". In: *Child Development* 62.5, p. 1066.

Wieland, Jan Willem (2016). "Responsibility for Strategic Ignorance". In: *Synthese* 194.11, pp. 4477–4497.

Zelditch, Morris (2001). "Thories of Legitmacy". In: *The Psychology of Legitimacy: Emerging Perspectives on Ideology, Justice, and Intergroup Relations*. Ed. by John T. Jost and Brenda Major. Cambridge University Press, pp. 33–53.

A. Appendix

A.1. Instructions

Pretest

English Version

Page 1:
Welcome
Dear Participants,
thank you for participating in this brief survey, which will take approximately 5 to 10 minutes to complete. Every twelfth participant will receive a $25 Amazon gift card at the end of the study. Your answers in this study will not affect the later drawing.
On the following page, we will first tell you what this study is about. Then we ask you for your opinion on the mechanisms presented. Finally, we will ask you for some general information about yourself.

Page 2:
What is it about?
Imagine the following situation:
A group of six people must be divided into three decision-makers and three recipients. Then the decision-makers have to divide a certain amount of money between themselves and the recipients. The following Figure A.1 is intended to illustrate the situation:

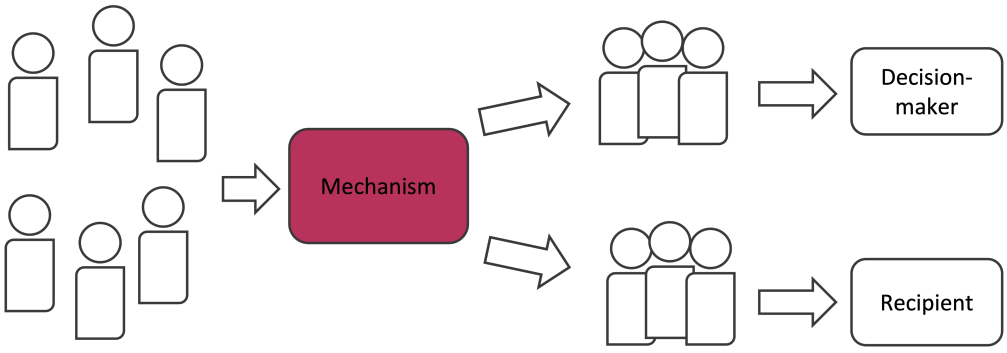


Figure A.1.: Schematic Procedure

On the following pages, four different group-building mechanisms are presented to you one after the other, each of which divides the six people into the two groups of decision-makers and recipients. We are interested in the extent to which the presented mechanisms lead to a justified division into decision-makers and recipients, i.e., to what extent the three decision-makers are legitimized to be allowed to make the subsequent decision. For this purpose, a scale is displayed on each page, on which we ask you to give your assessment.
Even though this is a hypothetical situation, please answer as truthfully as possible.

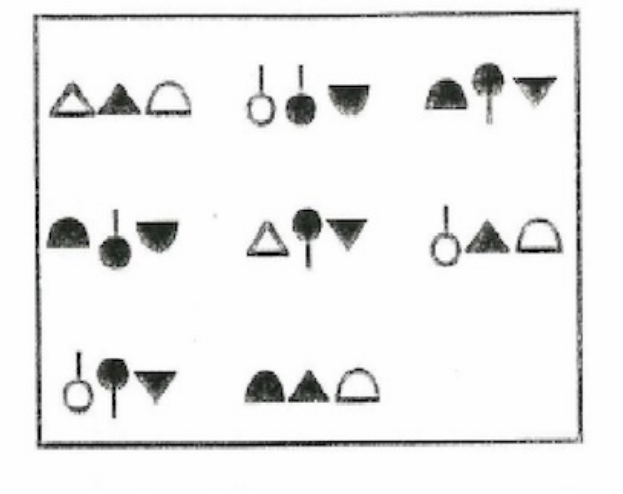
The four mechanisms appear in random order.

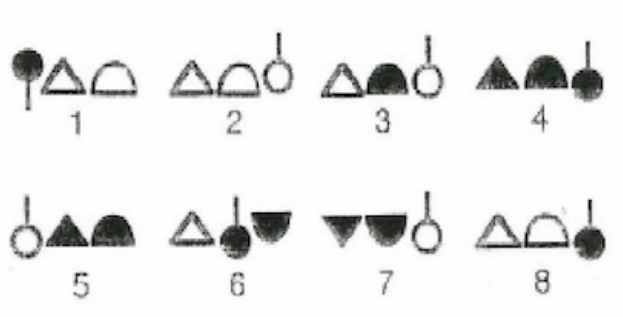
Page 3:
Mechanism 1 of 4: Final grade
Each participant reports his final grade. The three participants with the relatively better final grades are then assigned to the group of decision-makers, while the other three participants become recipients.
In your opinion, how justified is the final grade as a group-building mechanism?
A value of 1 means "not at all justified" and a value of 7 means "fully justified". You can use the values in between to grade your assessment.

Page 4:
Mechanism 2 of 4: Randomness
In random assignment, participants are randomly selected for the two groups of decision-makers and recipients. This means that each participant has an equal chance of being selected for one group or the other. Think of it as flipping a coin with a 50% chance of heads and a 50% chance of tails. Those for whom the coin came up tails, for example, are considered decision-makers, and those for whom the coin came up heads are considered recipients.
In your opinion, how justified is the random division into decision-makers and recipients?
A value of 1 means "not at all justified" and a value of 7 means "fully justified". You can use the values in between to grade your assessment.

Page 5:
Mechanism 3 of 4: Picture puzzle
The puzzle consists of 15 rounds. In each round, a picture with nine fields is presented, one of which is empty. Eight possible pictures are presented for this empty square, and the participant must choose the one that logically matches the other eight pictures. For each round, participants have one minute to make a choice.

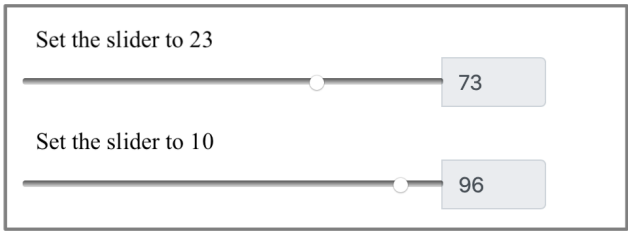
Here is an example of a round of the picture puzzle: Which of the following 8 pictures goes with the missing place in the box?





The participants are selected for the two groups according to the number of correctly solved picture problems. That is, the three participants who chose the most pictures correctly are assigned to the decision-makers, and the others are assigned to the recipients.
How justified do you think the picture puzzle is as a group-building mechanism?
A value of 1 means “not at all justified” and a value of 7 means “fully justified”. You can use the values in between to grade your assessment.

Page 6:
Mechanism 4 of 4: Slider
This mechanism consists of sliders that have values from 0 to 100 and are set to randomly selected numbers. Participants should use the mouse to move the slider to the specified position (23 or 10 in the example).
Here is an example of what the slider looks like:



Participants have a total of two minutes to set as many sliders correctly as possible. They are then divided into two groups, decision-makers and recipients, according to the number of sliders they set correctly. This means that the three participants who set the most sliders correctly are the decision-makers, and the other three are the recipients.
How justified do you think the slider is for dividing into decision-makers and recipients?
A value of 1 means “not at all justified” and a value of 7 means “fully justified”. You can use the values in between to grade your assessment.

Page 7:
Goodbye
Thank you for participating! You are now automatically entered into the Amazon gift card drawing and will be notified if you win.

German Version

Page 1:

Herzlich Willkommen

Liebe Teilnehmenden,

vielen Dank für Ihre Teilnahme an dieser kurzen Studie, welche ca. 5 bis 10 Minuten dauern wird. Jeder zwölfte Teilnehmende erhält bei Abschluss dieser Studie einen Amazon-Gutschein im Wert von 25 Euro. Ihre Antworten in dieser Studie haben dabei keinen Einfluss auf die spätere Verlosung.

Auf der folgenden Seite werden wir Ihnen zunächst vorstellen, worum es in dieser Studie geht. Anschließend bitten wir Sie, Ihre Einschätzung zu den dann vorgestellten Mechanismen abzugeben. Zum Schluss fragen wir Sie noch nach ein paar allgemeinen Angaben zu Ihrer Person.

Page 2:

Worum geht es?

Stellen Sie sich bitte folgende Situation vor:

Eine Gruppe von sechs Personen soll in drei Entscheider und drei Empfänger unterteilt werden. Die Entscheider dürfen anschließend einen gewissen Geldbetrag zwischen sich selbst und den Empfängern aufteilen. Die folgende Grafik A.2 soll Ihnen die Situation verdeutlichen:

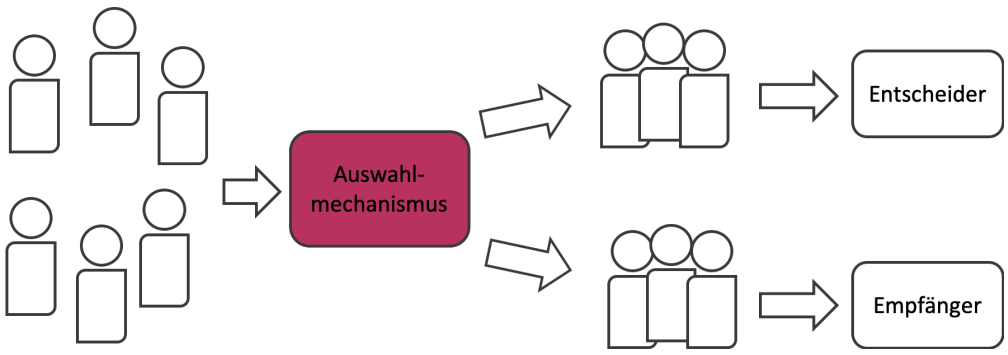


Figure A.2.: Schematischer Ablauf

Auf den folgenden Seiten werden Ihnen nacheinander vier verschiedene Auswahlmechanismen vorgestellt, die die sechs Personen jeweils auf unterschiedliche Art auf die zwei Gruppen Entscheider und Empfänger aufteilen. Für uns ist dabei von Interesse, inwieweit der dort jeweils vorgestellte Auswahlmechanismus zu einer gerechtfertigten Einteilung in Entscheider und Empfänger führt, also inwieweit die drei Entscheider legitimiert sind, die anschließende

Entscheidung treffen zu dürfen. Auf jeder Seite wird dafür eine Skala eingeblendet, auf der wir Sie bitten Ihre Einschätzung abzugeben.
Auch wenn es sich hierbei um eine hypothetische Situation handelt, bitten wir Sie, so wahrheitsgemäß wie möglich zu antworten.

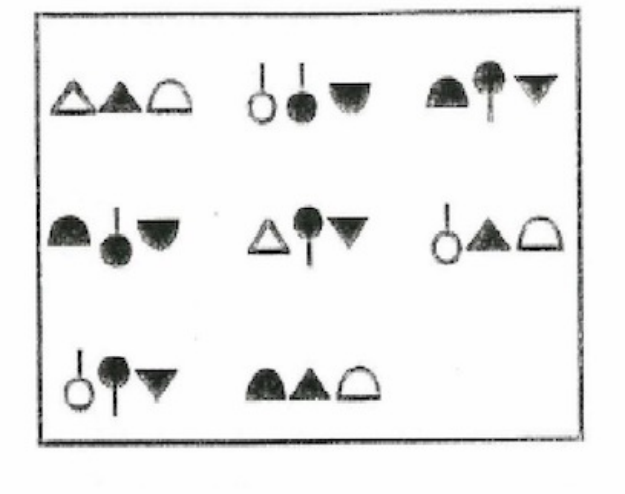
Die vier Mechanismen wurden in randomisierter Reihenfolge angezeigt.

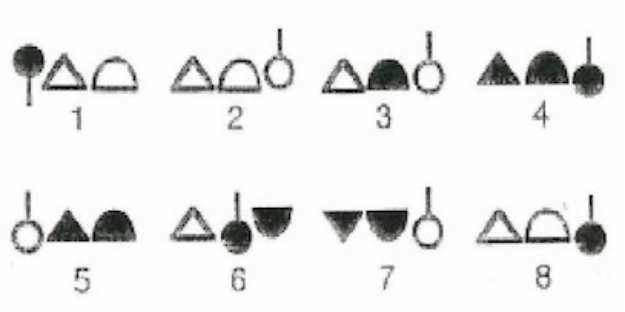
Page 3:
Mechanismus 1 von 4: Abitur- bzw. Abschlussnote
Jeder Teilnehmende gibt seine Abitur- bzw. Abschlussnote an. Die drei Teilnehmenden mit den relativ besseren Abschlussnoten werden dann der Gruppe der Entscheider zugeordnet, die anderen drei Teilnehmenden werden Empfänger.
Wie gerechtfertigt ist Ihrer Meinung nach die Abitur- bzw. Abschlussnote als Auswahlmechanismus?
Der Wert 1 bedeutet "ganz und gar nicht gerechtfertigt" und der Wert 7 "voll und ganz gerechtfertigt". Mit den Werten dazwischen können Sie Ihre Einschätzung abstufen.

Page 4:
Mechanismus 2 von 4: Zufall
Bei der zufälligen Einteilung werden die Teilnehmenden per Zufallsentscheid für die beiden Gruppen Entscheider und Empfänger ausgewählt. Das heißt, jeder Teilnehmende hat die gleiche Wahrscheinlichkeit für die eine oder andere Gruppe ausgewählt zu werden. Stellen Sie sich das wie das Werfen einer Münze vor, bei der mit 50% Wahrscheinlichkeit Kopf und mit 50% Wahrscheinlichkeit Zahl erscheint. Diejenigen bei denen die Münze dann beispielsweise Zahl angezeigt hat, werden den Entscheidern und diejenigen mit Kopf den Empfängern zugeordnet.
Wie gerechtfertigt ist Ihrer Meinung nach die zufällige Einteilung in Entscheider und Empfänger?
Der Wert 1 bedeutet "ganz und gar nicht gerechtfertigt" und der Wert 7 "voll und ganz gerechtfertigt". Mit den Werten dazwischen können Sie Ihre Einschätzung abstufen.

Page 5:
Mechanismus 3 von 4: Bilderrätsel
Das Bilderrätsel besteht aus 15 Runden. Dabei wird in jeder Runde ein Bild mit neun Feldern präsentiert, wovon eines leer ist. Für dieses leere Feld stehen acht mögliche Bilder zur Verfügung, aus denen dasjenige ausgewählt werden soll, welches logisch zu den anderen

acht Bildern passt. Für jede Runde bekommen die Teilnehmenden eine Minute Zeit, um eine Wahl zu treffen.
Hier ein Beispiel wie eine Runde des Bilderrätsels aussieht: Welches der folgenden 8 Bilder gehört an die fehlende Stelle im Kasten?

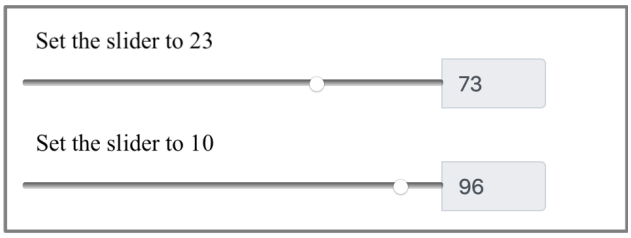




Die Teilnehmenden werden anschließend nach Anzahl der richtig gelösten Bilderaufgaben für die zwei Gruppen ausgewählt. Das heißt, die drei Teilnehmenden, die relativ am meisten Bildern richtig erkannt haben, werden den Entscheidern zugeordnet, die anderen den Empfängern.
Wie gerechtfertigt ist Ihrer Meinung nach das Bilderrätsel als Auswahlmechanismus?
Der Wert 1 bedeutet “ganz und gar nicht gerechtfertigt” und der Wert 7 “voll und ganz gerechtfertigt”. Mit den Werten dazwischen können Sie Ihre Einschätzung abstufen.

Page 6:
Mechanismus 4 von 4: Schieberegler

Bei diesem Mechanismus ist ein Schieberegler vorgegeben, welcher Werte von 0 bis 100 aufweist und auf einer zufällig eingestellten Zahl steht. Die Teilnehmenden sollen dann mit der Maus die angegebene Position des Schiebereglers (im Beispiel 23 bzw. 10) einstellen.
Hier ein Beispiel wie der Schieberegler aussieht:



Insgesamt stehen den Teilnehmenden zwei Minuten Zeit zur Verfügung, um so viele Schieberegler wie möglich richtig einzustellen. Die Teilnehmenden werden dann nach der Anzahl der richtig eingestellten Schieberegler für die zwei Gruppen Entscheider und Empfänger ausgewählt. Das heißt, die drei Teilnehmenden, die am meisten Schieberegler richtig eingestellt haben, sind Entscheider, die anderen drei Empfänger.
Wie gerechtfertigt ist Ihrer Meinung nach der Schieberegler zur Einteilung in Entscheider und Empfänger?
Der Wert 1 bedeutet “ganz und gar nicht gerechtfertigt” und der Wert 7 “voll und ganz gerechtfertigt”. Mit den Werten dazwischen können Sie Ihre Einschätzung abstufen.

Page 7:
Verabschiedung
Vielen Dank für Ihre Teilnahme! Sie nehmen nun automatisch an der Verlosung der Amazon-Gutscheine teil und werden bei einem Gewinn informiert.

Main experiment

English Version

General explanations for the participants

Welcome to the experiment. If you read the instructions carefully and follow the rules, you can earn money in this experiment. You will receive **6 euros as a fixed payout**. Depending on your decisions and the decisions of the other participants, you can earn **additional money**. The money will be paid to you in cash immediately afterwards. During the experiment we do not speak of euros, but of points. These are converted according to the following exchange rate:

1 point = 1 euro

Talking to other participants is not allowed during the entire experiment. If you have any questions, please direct them only to us. We will be happy to answer your questions individually. It is very important that you follow this rule. Otherwise, the results of this experiment will be scientifically worthless.

This experiment consists of **2 parts**, both parts will be explained one after the other. The experiment should take 60 to 75 minutes to complete. The first part of the experiment is explained below.

Detailed information about the first part of the experiment

The first part of the experiment is about the distribution of 10 points between two participants A and B. While participant A decides about the distribution of the points, participant B has the possibility to take points away from participant A.

You will first take the role of participant A and then the role of participant B. This means that you will first have the opportunity to determine the distribution of points and then to distribute the deduction points.

Your task as "Participant A":

First, you are participant A and you choose how the 10 points will be divided between you and a random other participant B. You have to choose between two possible allocations: allocation 1 (5; 5), where you (participant A) and participant B each get 5 points, and allocation 2 (9; 1), where you (participant A) get 9 points and participant B gets 1 point. In Figure A.3 you see your decision screen as participant A.

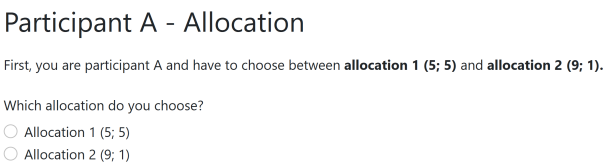


Figure A.3.: Decision Screen as Participant A

Your task as "Participant B":
Afterwards you decide as participant B. Since you are now in the position of participant B, you can assign deduction points to participant A. In Figure A.4 you can see the two possible allocations 1 and 2, where you can assign up to 5 deduction points to participant A for each allocation.

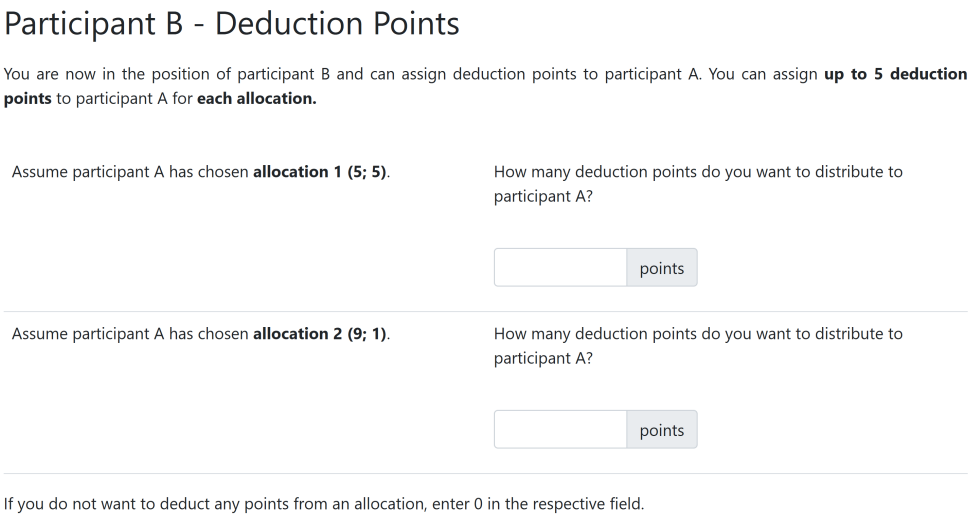


Figure A.4.: Decision Screen as Participant B

Deducting points has a **cost**. If you want to deduct points from participant A, you must give up 1 point to be able to deduct up to 5 points. **Any integer number of points between 0 and 5 points** can be deducted for both possible choices. Once at least 1 point is subtracted, the cost is 1 point. So the cost of subtracting points is always 1 point, regardless of the number of points subtracted.

Your payout from part 1:
After you have made your decision as participant A and participant B, you will be randomly assigned to another participant in the experiment. Both of you will then be randomly as-

signed a role, either participant A or participant B, and your decision in that role will be selected as **relevant for payout**.

Your payout as "Participant A"
If you are selected as participant A, you will receive the number of points you kept for yourself minus the number of points deducted by your assigned participant B. For example, if you selected allocation 1 (5; 5) and participant B deducted 2 points from your allocation, you would receive 5 points - 2 points = 3 points.

Your payout as "Participant B"
If you are selected as participant B, you will receive the number of points that participant A has assigned to you, minus 1 point if you have assigned deductions for that assignment. That is, if contestant A chose allocation 1 (5; 5) and you gave away 2 points for that allocation, you will receive 5 points - 1 point = 4 points. If you did not deduct any points for allocation 1 (5; 5), you do not incur a cost of 1 point, regardless of whether and how many points you deducted for allocation 2 (9; 1).

You will know your payout from this part of the experiment at the end of the entire experiment. Before we proceed with the second part of the experiment, we will ask you to give us your high school graduation grade. Then we will give you the instructions for the second part of the experiment.

If you have any questions, raise your hand outside the cabin and wait quietly until someone comes to you.
Thank you for your participation and have fun with the experiment.

Instructions for the second part of the experiment

Your payout from the first part of the experiment is now fixed and will be sent to you at the end of the experiment. The following second part of the experiment has no effect on your payout from the first part. You will now receive the **instructions for the second part of the experiment.**

Role assignment in part 2:
At the beginning of the second part, you will be assigned to a role of participant A or participant B *randomly* [based on your final grade. Those with the relatively higher grades will be assigned to participant A and the others to participant B.] This role assignment will be maintained throughout the second part of the experiment.

Part 2 of the experiment consists of a total of 6 rounds. At the beginning of each new round, you will be randomly and anonymously divided into groups with other participants A and B who are also taking part in this experiment. Each group consists of three participants A (A1, A2, and A3) and three participants B (B1, B2, and B3). Thus, you will always be paired with two other participants in your role and three other participants in the other role. All participants remain anonymous, as do you. After each round, the groups are rearranged, with each participant keeping his or her role A or B. However, you will be grouped with different participants A and B than in the previous round.

The decision-making situation:
The second part of the experiment consists of two stages. In **stage 1**, the three participants A of a group decide **sequentially** how to divide 30 points between the three participants A and the three participants B within their group. Participants A must choose between two possible allocations of points:

- **Allocation 1 (5, 5, 5; 5, 5, 5):** Participant A and participant B receive 5 points each.

- **Allocation 2 (9, 9, 9; 1, 1, 1):** The three participants A receive 9 points each and the three participants B receive 1 point each.

The allocation that receives the majority of votes from participant A will be implemented. Then, in **stage 2**, participants B will have the opportunity to assign deduction points to the three participants A.

Stage 1 - Allocation decision:
At the beginning of each new round, participants A learn the position (A1, A2 or A3) to which they have been assigned for that round. It is possible to get the same position several times, even consecutively. However, the group composition changes with each new round. The participants A vote **sequentially** on the allocation, whereby the allocation *1 (5, 5, 5; 5, 5, 5)* [2 (9, 9, 9; 1, 1, 1)] is already preselected. Each participant A then has 30 seconds to choose one of the two allocations.

The decision as "Participant A1":
The first participant to make a decision is participant A1, who has 30 seconds to select an allocation and click the 'next button'. If contestant A1 has not clicked the 'next button' before the time expires, the marked allocation will be scored.

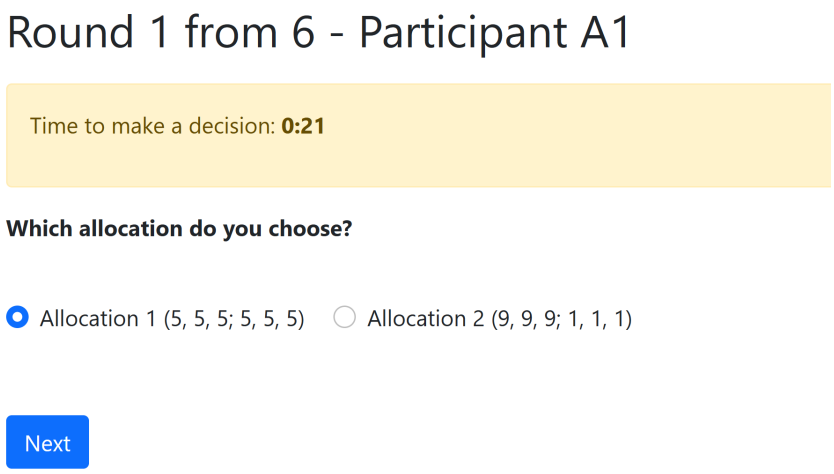


Figure A.5.: Decision Screen of Participant A1

The decision as "Participant A2":
Participant A, who makes the second decision, is participant A2. Participant A2 is shown how participant A1 decided before making his own decision. Again, an allocation is preselected and participant A2's decision screen is displayed for 30 seconds.

The decision as "Participant A3":
Participant A, who makes the third decision, becomes participant A3. Participant A3 is shown how participants A1 and A2 have decided before being given 30 seconds to make his own decision.

The distribution for which **at least two of the three participants A decide** is implemented for that round. The voting result is therefore fixed as soon as two participants A have chosen the same distribution.

The decision as "Participant B":
While participants A make their decisions one by one, we would like to know from participants B which division they think participants A1, A2 and A3 will choose. For this purpose, participants B are assigned to one of the three positions B1, B2 and B3 in each new round. They then go through the decision situation one by one, just like participants A, and have 30 seconds to state their expected distribution. This means that participant B1 states his expectations first, followed by participant B2, who is shown the expectation stated by participant B1. After that, participant B3 states his expected distribution, while being shown the expectations of participants B1 and B2. (see Figure A.6).

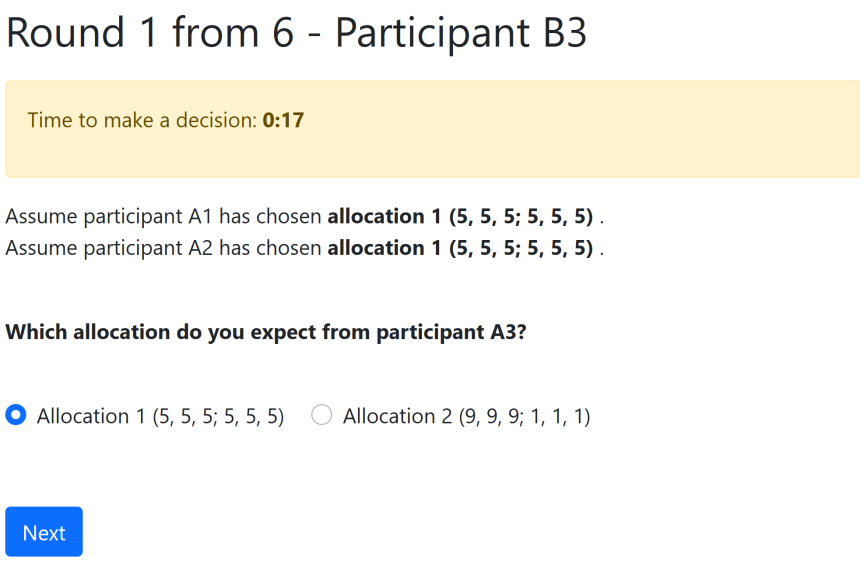


Figure A.6.: Decision Screen of Participant B3

Stage 2 – Distribution of deduction points:
After participants A have determined their allocation and participants B have stated their

expectations, the next step is for participants B to distribute their deduction points.

The decisions of participants B:
Participant B learns the allocation that has been made and how each participant A has decided. **Each participant B** then has the opportunity to deduct points from the payouts of participants A1, A2 and A3.
Deducting points has a **cost**, as in part 1 of the experiment: If participant B wants to deduct points from participant A, he must give up 1 point in order to be able to deduct up to 7 points from participant A in his group. **Any integer number of points between 0 and 7 points can be deducted.** Once at least 1 point has been deducted, the participant B who deducted the points has a cost of 1 point. The cost of deducting points is therefore always 1 point, regardless of the number of points deducted. For example, if participant B wants to deduct 7 points from participant A3, participant A3's payout is reduced by **7 points** and participant B's payout is reduced by **1 point**.

The only restriction on the deduction of points is that no more than a total of 7 points can be deducted and that no more points can be deducted from a participant A than he has received in the implemented allocation. If the allocation is 1 (5, 5, 5; 5, 5, 5), a total of up to 7 points can be deducted, but no participant A can be deducted more than 5 points.

At the end of the six rounds, one round is randomly selected as relevant to the payout. In each group, **a participant B is then randomly selected** and only the deduction points of this one randomly selected participant B are converted. In each group, therefore, **only this one participant B incurs a cost of 1 point in this one round** (if he deducts points).
Since you do not know which round will be selected for payout and whether you will be drawn, the deduction points you distribute in each of the 6 rounds may determine the payouts at the end of the experiment.

The decision as "Participant A":
While participants B are assigning their deduction points, we want to know from participants A how many deduction points they expect to receive. Thus, as participant A, you will see which allocation was chosen by participants A1, A2, A3 within your group, and you must indicate how many deduction points you think participant A will receive for the decision he

made.

Your payoff from part 2 of the experiment:
After the 6th round you will see which round and which anonymous participant B was chosen as relevant for the payout. You will also see which distribution was chosen by the majority in the payout relevant round and if and how many deduction points were distributed by the randomly chosen participant B.

Your payout as "Participant A":
As participant A, your payout will be equal to your allocation minus the deduction points of the randomly selected participant B. For example, if the majority of your group chose allocation 1 (5, 5, 5; 5, 5, 5), and the selected participant B distributed 1 deduction point to each participant A, you as participant A will receive **5 points minus 1 deduction point, so 4 points** from part 2.

Your payout as "Participant B":
As a participant B, your payout will be equal to your assigned allocation minus one point if you are selected to distribute deductions.
If the majority of participants A in your group chose allocation 1 (5, 5, 5; 5, 5, 5), you will receive 5 points. If you are also participant B, who is allowed to assign deduction points, and you have done so, you will be deducted 1 point. So you get **5 points minus 1 point (cost of the deduction points), i.e. 4 points** from part 2.

After the second part, the actual experiment is over. We will ask you to answer the following questionnaire. You will then receive your payout. Please remain seated until we call your cubicle number.

If you have any questions, please raise your hand outside the booth and wait quietly until someone comes to you. If you have no questions, you will be given the password to start the second part of the experiment.

German Version

Instruktionen Teil 1

Allgemeine Erklärungen für die Teilnehmer
Willkommen zum Experiment. Wenn Sie die Instruktionen aufmerksam lesen und alle Regeln beachten, können Sie in diesem Experiment Geld verdienen. Sie erhalten **6 Euro als feste Auszahlung**. Abhängig von Ihren Entscheidungen und den Entscheidungen anderer Teilnehmer können Sie **zusätzlich Geld verdienen**. Das Geld wird im Anschluss sofort in bar an Sie ausbezahlt. Während des Experiments sprechen wir nicht von Euro, sondern von Punkten. Diese werden gemäß folgendem Wechselkurs umgerechnet:

1 Punkt = 1 Euro

Während des gesamten Experiments ist das Sprechen mit anderen Teilnehmern nicht erlaubt. Wenn Sie Fragen haben, richten Sie diese bitte ausschließlich an uns. Wir beantworten Ihre Fragen gerne individuell. Die Einhaltung dieser Regel ist sehr wichtig. Andernfalls sind die Ergebnisse dieses Experiments wissenschaftlich wertlos.
Dieses Experiment besteht aus **2 Teilen**, wobei beide Teile nacheinander einzeln erläutert werden. Das Experiment wird voraussichtlich 60 bis 75 Minuten dauern. Im Folgenden wird Ihnen nun der erste Teil des Experiments erläutert.

Detaillierte Informationen zum 1. Teil des Experiments
Im ersten Teil des Experiments geht es um die Aufteilung von 10 Punkten zwischen zwei Teilnehmern A und B. Teilnehmer A entscheidet dabei über die Aufteilung der Punkte, Teilnehmer B hat daran anschließend die Möglichkeit Teilnehmer A Punkte abzuziehen.
Sie werden zunächst die Rolle von Teilnehmer A einnehmen und danach die Rolle von Teilnehmer B. Das heißt, Sie haben zuerst die Möglichkeit die Aufteilung der Punkte zu bestimmen und anschließend Abzugspunkte zu verteilen.

Ihre Aufgabe als „Teilnehmer A“:
Zunächst sollen Sie als Teilnehmer A wählen, wie die 10 Punkte zwischen Ihnen und einem zufälligen anderen Teilnehmer B aufgeteilt werden sollen. Dabei haben Sie die Wahl zwischen zwei möglichen Aufteilungen: Aufteilung 1 (5; 5), bei der Sie selbst und der andere Teilnehmer B jeweils **5 Punkte** erhalten, und Aufteilung 2 (9; 1), bei der Sie selbst **9 Punkte** erhalten und der andere Teilnehmer B **1 Punkt**. In Abbildung A.7 sehen Sie Ihren Entscheidungsbildschirm als Teilnehmer A.

Figure A.7.: Entscheidungsbildschirm als Teilnehmer A

Ihre Aufgabe als „Teilnehmer B“:
Anschließend treffen Sie eine Entscheidung als Teilnehmer B. Dabei dürfen Sie Abzugspunkte an Teilnehmer A für beide mögliche Aufteilungen verteilen. Ihnen stehen hierfür jeweils bis zu 5 Abzugspunkte zur Verfügung. Das heißt, Sie können sowohl bei Aufteilung 1 wie auch bei Aufteilung 2 bis zu 5 Abzugspunkte an Teilnehmer A verteilen (siehe Abbildung A.8).

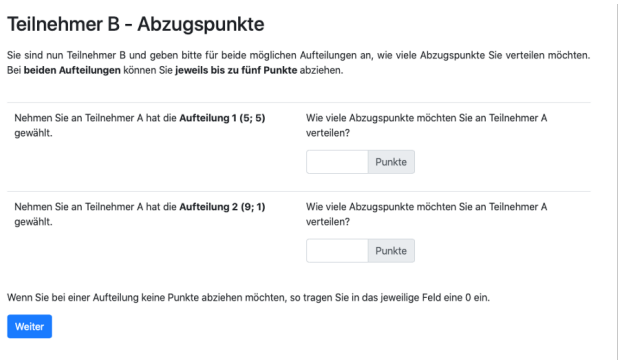


Figure A.8.: Entscheidungsbildschirm als Teilnehmer B

Das Abziehen von Punkten ist mit **Kosten** verbunden: Wenn Sie dem Teilnehmer A Punkte abziehen möchte, müssen Sie dafür 1 Punkt aufgeben, um bis zu 5 Punkte abziehen zu können. Es kann bei beiden möglichen Aufteilungen **jede beliebige ganze Punktzahl zwischen 0 und 5 Punkten** abgezogen werden. Sobald mindestens 1 Punkt abgezogen wird, entstehen die Kosten in Höhe von 1 Punkt. Die Kosten, um Punkte abzuziehen, betragen also immer 1 Punkt, unabhängig von der Anzahl abgezogener Punkte.

Ihre Auszahlung aus Teil 1:
Nachdem Sie Ihre Entscheidung als Teilnehmer A und B getroffen haben, wird Ihnen ein anderer Teilnehmer des Experiments zufällig zugeordnet. Ihnen beiden wird dann eine Rolle, entweder Teilnehmer A oder Teilnehmer B, zufällig zugeordnet und Ihre in der entsprechenden Rolle getroffene Entscheidung als **auszahlungsrelevant** ausgewählt.

Ihre Auszahlung als „Teilnehmer A“

Sollten Sie als Teilnehmer A ausgewählt werden, erhalten Sie die Anzahl an Punkten, die Sie für sich selbst behalten haben, abzüglich der Abzugspunkte des Ihnen zugeteilten Teilnehmer B. Das heißt, sollten Sie die Aufteilung 1 (5; 5) gewählt haben und der Ihnen zugeloste Teilnehmer B 2 Abzugspunkte für diese Aufteilung verteilt haben, erhalten Sie 5 Punkte – 2 Punkte = 3 Punkte.

Ihre Auszahlung als „Teilnehmer B“

Sollten Sie als Teilnehmer B ausgewählt werden, erhalten Sie die Anzahl an Punkten, die Ihnen Teilnehmer A zugeteilt hat, abzüglich 1 Punkt, wenn Sie für die entsprechende Aufteilung Abzugspunkte verteilt haben. Das heißt, sollte Teilnehmer A die Aufteilung 1 (5; 5) gewählt haben und Sie für diese Aufteilung 2 Abzugspunkte verteilt haben, erhalten Sie 5 Punkte - 1 Punkt = 4 Punkte. Sollten Sie bei der Aufteilung 1 (5; 5) keine Abzugspunkte verteilt haben, fallen auch keine Kosten von 1 Punkt an, unabhängig davon, ob und wie viele Punkte Sie bei der Aufteilung 2 (9; 1) abgezogen haben.

Ihre Auszahlung aus diesem Teil des Experiments erfahren Sie am Ende des gesamten Experiments. Bevor es dann mit dem zweiten Teil des Experiments weiter geht, bitten wir Sie Ihre Abitur- bzw. Abschlussnote, also die Note Ihres höchsten allgemeinbildenden Schul-abschlusses, anzugeben. Anschließend teilen wir Ihnen dann die Instruktionen für den zweiten Teil des Experiments aus.

Falls Sie nun noch Fragen haben, heben Sie die Hand aus der Kabine und warten Sie ruhig, bis jemand zu Ihnen kommt.
Vielen Dank schon mal für Ihre Teilnahme und viel Spaß im Experiment.

Instruktionen zum zweiten Teil des Experiments

Ihre Auszahlung aus dem ersten Teil des Experiments steht nun fest und wird Ihnen am Ende des Experiments mitgeteilt. Der folgende zweite Teil des Experiments hat keine Auswirkung auf Ihre Auszahlung aus dem ersten Teil. Sie erhalten nun die **Instruktionen zum zweiten Teil des Experiments.**

Rolleneinteilung in Teil 2:
Zu Beginn des zweiten Teils werden Sie *zufällig* [aufgrund Ihrer Abitur- bzw. Abschlussnote] einer Rolle Teilnehmer A oder Teilnehmer B zugeordnet. [Diejenigen, mit den relativ besseren Noten, werden den Teilnehmern A und die anderen den Teilnehmern B zugeordnet.] Diese Rolleneinteilung bleibt dann im kompletten zweiten Teil des Experiments bestehen und ist unabhängig der Rolle, die Ihnen im ersten Teil des Experiments zugeordnet wurde.
Teil 2 des Experiments wird über insgesamt 6 Runden gespielt. Zu Beginn jeder neuen Runde werden Sie mit anderen Teilnehmern A und B, die auch an diesem Experiment teilnehmen, zufällig und anonym in Gruppen eingeteilt. Jede Gruppe besteht dabei aus drei Teilnehmern A (A1, A2 und A3) und drei Teilnehmern B (B1, B2 und B3). Ihnen sind somit immer zwei andere Teilnehmer Ihrer Rolle und drei Teilnehmer der anderen Rolle zugeordnet. Alle Teilnehmer bleiben, wie Sie auch, anonym. Nach jeder Runde werden die Gruppen neu zusammengesetzt, wobei jeder Teilnehmer seine Rolle A oder B beibehält. Sie werden allerdings mit anderen Teilnehmern A und B gruppiert als in der vorherigen Runde.

Die Entscheidungssituation:
Der zweite Teil des Experiments besteht aus zwei Stufen. In **Stufe 1** entscheiden die drei Teilnehmer A einer Gruppe **nacheinander**, wie 30 Punkte zwischen den drei Teilnehmern A und den drei Teilnehmern B innerhalb Ihrer Gruppe aufgeteilt werden sollen. Die Teilnehmer A müssen sich dabei zwischen zwei möglichen Aufteilungen entscheiden:

- **Aufteilung 1 (5, 5, 5; 5, 5, 5):** Die Teilnehmer A und die Teilnehmer B bekommen jeweils 5 Punkte.

- **Aufteilung 2 (9, 9, 9; 1, 1, 1):** Die drei Teilnehmer A bekommen jeweils 9 Punkte und die drei Teilnehmer B bekommen jeweils 1 Punkt.

Die Aufteilung, welche die Mehrheit der Stimmen der Teilnehmer A bekommt, wird umgesetzt. Anschließend haben die Teilnehmer B in **Stufe 2** die Möglichkeit Abzugspunkte an die drei Teilnehmer A zu verteilen.

Stufe 1 - Entscheidung über die Aufteilung:
Zu Beginn jeder neuen Runde erfahren die Teilnehmer A zunächst welche Position (A1, A2 oder A3) sie in der jeweiligen Runde zugeteilt bekommen haben. Dabei ist es möglich dieselbe Position mehrmals, auch nacheinander, zu bekommen. Die Gruppenzusammensetzung ändert sich jedoch mit jeder neuen Runde.
Die Teilnehmer A stimmen **nacheinander** über die Aufteilung ab, wobei Aufteilung *1 (5, 5, 5; 5, 5, 5)* [2 (9, 9, 9; 1, 1, 1)] voreingestellt, also bereits ausgewählt, ist. Die Teilnehmer A haben dann jeweils 30 Sekunden Zeit, um sich für eine der beiden Aufteilungen zu entscheiden.

Die Entscheidung als „Teilnehmer A1“:
Der Teilnehmer A, der zuerst entscheidet, ist Teilnehmer A1 und hat 30 Sekunden Zeit, eine Aufteilung zu wählen und auf „weiter“ zu klicken. Sollte Teilnehmer A1 vor Ablauf der Zeit nicht „weiter“ geklickt haben, wird die dann markierte Aufteilung gewertet.

Figure A.9.: Entscheidungsbildschirm des Teilnehmers A1

Die Entscheidung als „Teilnehmer A2“:
Der Teilnehmer A, der als zweites entscheidet, ist Teilnehmer A2. Teilnehmer A2 bekommt angezeigt, wie sich Teilnehmer A1 entschieden hat, bevor er seine eigene Entscheidung trifft. Auch hier ist wieder eine Aufteilung voreingestellt und der Entscheidungsbildschirm von Teilnehmer A2 wird für 30 Sekunden angezeigt.

Die Entscheidung als „Teilnehmer A3“:

Der Teilnehmer A, der als drittes entscheidet, ist Teilnehmer A3. Teilnehmer A3 bekommt dabei angezeigt, wie sich Teilnehmer A1 und A2 entschieden haben, bevor er 30 Sekunden Zeit hat seine eigene Entscheidung zutreffen.

Die Aufteilung, für die **sich mindestens zwei der drei Teilnehmer A entscheiden**, wird für die jeweilige Runde umgesetzt. Das Abstimmungsergebnis steht also fest, sobald sich zwei Teilnehmer A für dieselbe Aufteilung entschieden haben.

Die Entscheidung als „Teilnehmer B“:
Während die Teilnehmer A nacheinander ihre Entscheidung treffen, möchten wir von den Teilnehmern B wissen, welche Aufteilung Ihrer Meinung nach jeweils von den Teilnehmern A1, A2 und A3 gewählt wird. Dafür werden den Teilnehmern B in jeder neuen Runde eine der drei Positionen B1, B2 und B3 zugelost. Sie durchlaufen dann ebenso wie die Teilnehmer A nacheinander die Entscheidungssituation und haben jeweils 30 Sekunden Zeit, ihre erwartete Aufteilung anzugeben. Das heißt, Teilnehmer B1 gibt als erstes seine Erwartungen an, dann folgt Teilnehmer B2, welcher angezeigt bekommt, welche Erwartung Teilnehmer B1 angegeben hat. Anschließend gibt Teilnehmer B3 an, welche Aufteilung er erwartet, wobei er die Erwartungen des Teilnehmers B1 und B2 angezeigt bekommt (siehe Abbildung A.10).

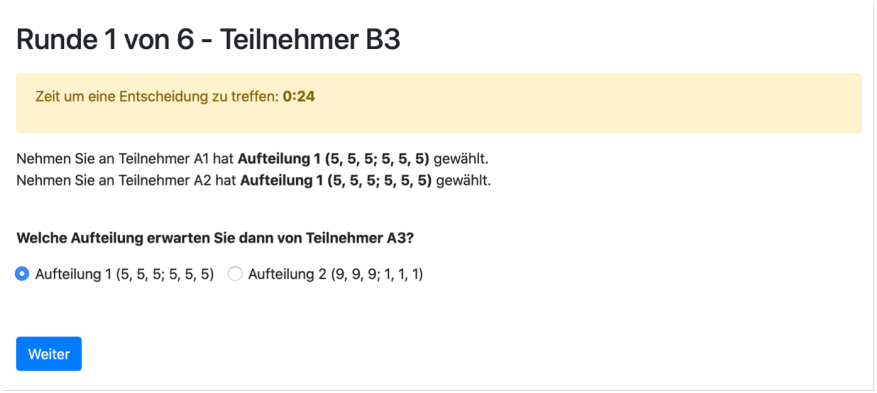


Figure A.10.: Entscheidungsbildschirm von Teilnehmer B3

Stufe 2 - Verteilen von Abzugsunkten:
Nachdem die Aufteilung von den Teilnehmern A bestimmt wurde und die Teilnehmer B Ihre Erwartungen angegeben haben, folgt im nächsten Schritt das Verteilen von Abzugspunkten durch die Teilnehmer B.

Die Entscheidungen der Teilnehmer B:

Die Teilnehmer B erfahren das Ergebnis der Abstimmung und wie sich jeder einzelne Teilnehmer A entschieden hat. **Jeder Teilnehmer B** hat dann die Möglichkeit, den Teilnehmern A1, A2 und A3 Punkte von deren Auszahlung abzuziehen (siehe Abbildung A.11).

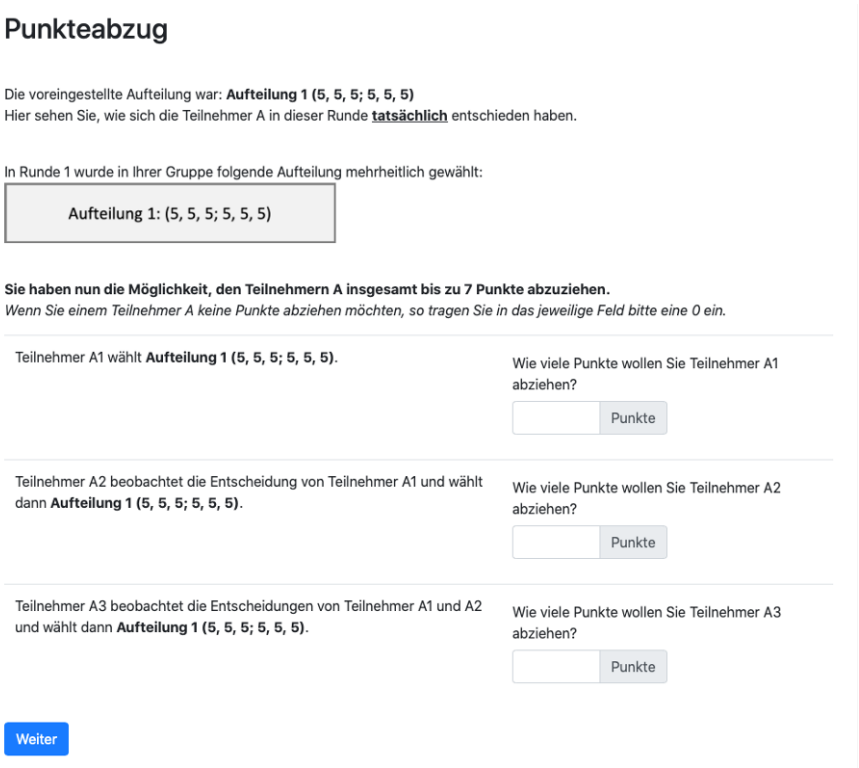


Figure A.11.: Entscheidungsbildschirm als Teilnehmer B – Punkteabzug

Das Abziehen von Punkten ist, wie in Teil 1 des Experiments, mit **Kosten** verbunden: Wenn der Teilnehmer B den Teilnehmern A Punkte abziehen möchte, muss er 1 Punkt aufgeben, um den Teilnehmern A seiner Gruppe bis zu 7 Punkte abziehen zu können. Es kann **jede beliebige ganze Punktzahl zwischen 0 und 7 Punkten** abgezogen werden. Sobald mindestens einem Teilnehmer A mindestens 1 Punkt abgezogen wird, hat der Teilnehmer B, der Punkte abzieht, Kosten in Höhe von 1 Punkt. Die Kosten, um Punkte abzuziehen betragen also immer 1 Punkt, unabhängig von der Anzahl abgezogener Punkte. Wenn ein Teilnehmer B dem Teilnehmer A3 beispielsweise 7 Punkte abziehen möchte, so reduziert sich die Auszahlung des Teilnehmers A3 um **7 Punkte** und die Auszahlung des Teilnehmers B um **1 Punkt**.
Die einzigen Beschränkungen beim Punkteabzug sind, dass nie mehr als insgesamt 7 Punkte abgezogen werden können und dass einem Teilnehmer A nie mehr Punkte abgezogen werden können, als dieser bei der mehrheitlich gewählten

Aufteilung bekommen hat. Wenn also die Aufteilung 1 (5, 5, 5; 5, 5, 5) resultiert, dann können insgesamt bis zu 7 Punkte, jedoch keinem Teilnehmer A mehr als 5 Punkte abgezogen werden.

Am Ende der sechs Runden wird eine zufällige Runde als auszahlungsrelevant ausgewählt. In jeder Gruppe wird dann **ein Teilnehmer B zufällig ausgewählt** und nur die Abzugspunkte dieses einen zufällig ausgewählten Teilnehmers B werden dann umgesetzt. In jeder Gruppe entstehen also **nur diesem einen Teilnehmer B in dieser einen Runde Kosten in Höhe von 1 Punkt** (sofern er Punkte abzieht).
Da Sie nicht wissen, welche Runde als auszahlungsrelevant ausgewählt wird und ob Sie ausgelost werden, können Ihre verteilten Abzugspunkte in jeder der 6 Runden für die Auszahlungen am Ende des Experiments bestimmend sein.

Die Entscheidung als „Teilnehmer A“:
Während die Teilnehmer B ihre Abzugspunkte verteilen, möchten wir von den Teilnehmern A wissen, wie viele Abzugspunkte sie erwarten zu bekommen. Das heißt, als Teilnehmer A sehen Sie, welche Aufteilung von den Teilnehmern A1, A2, A3 innerhalb Ihrer Gruppe jeweils gewählt wurden und haben dann die Möglichkeit anzugeben, wie viele Abzugspunkte, Ihrer Meinung nach, der jeweilige Teilnehmer A für die jeweils getroffene Entscheidung erhält.

Ihre Auszahlung aus Teil 2 des Experiments:
Nach der 6. Runde sehen Sie, welche Runde und welcher anonyme Teilnehmer B als auszahlungsrelevant ausgewählt wurde. Außerdem wird Ihnen angezeigt, welche Aufteilung in der auszahlungsrelevanten Runde mehrheitlich gewählt wurde, und ob und wie viele Abzugspunkte vom zufällig ausgewählten Teilnehmer B verteilt wurden.

Auszahlung als „Teilnehmer A“:
Als Teilnehmer A setzt sich Ihre Auszahlung aus der mehrheitlich gewählten Aufteilung, abzüglich der Abzugspunkte des zufällig ausgewählten Teilnehmers B, zusammen. Sollten Sie in Ihrer Gruppe beispielsweise mehrheitlich Aufteilung 1 (5, 5, 5; 5, 5, 5) gewählt haben und der ausgewählte Teilnehmer B an jeden Teilnehmer A 1 Abzugspunkt verteilt haben, erhalten Sie als Teilnehmer A **5 Punkte abzüglich 1 Abzugspunkt, also 4 Punkte** aus Teil 2.

Auszahlung als „Teilnehmer B“:

Als Teilnehmer B setzt sich Ihre Auszahlung aus der mehrheitlich gewählten Aufteilung, abzüglich eines Punktes, falls Sie zur Verteilung von Abzugspunkten ausgewählt wurden, zusammen.
Sollten sich die Teilnehmer A in Ihrer Gruppe mehrheitlich für Aufteilung 1 (5, 5, 5; 5, 5, 5) entschieden haben, erhalten Sie 5 Punkte. Wenn Sie weiterhin derjenige Teilnehmer B sind, der Abzugspunkte verteilen darf und dieses auch gemacht haben, wird Ihnen zusätzlich 1 Punkt abgezogen. Sie bekommen also **5 Punkte abzüglich 1 Punkt (Kosten für die Abzugspunkte), also 4 Punkte** aus Teil 2.

Nach diesem zweiten Teil ist das eigentliche Experiment zu Ende. Wir bitten Sie den anschließenden Fragebogen zu beantworten. Danach erhalten Sie Ihre Auszahlung. Bitte bleiben Sie sitzen, bis wir Ihre Kabinennummer aufrufen.

Falls Sie nun noch Fragen haben, heben Sie die Hand aus der Kabine und warten Sie ruhig, bis jemand zu Ihnen kommt. Wenn Sie keine Fragen haben erhalten Sie nun das Passwort, um den zweiten Teil des Experiments zu starten.

A.2. Pretest

Table A.1.: Margins of tobit Regression - Pretest

Model	Mechanism	Margins	Random		IQ		Slider	
			χ^2	$p(\chi^2)$	χ^2	$p(\chi^2)$	χ^2	$p(\chi^2)$
					Overall			
	Random	4.882***		-		-		-
(2)	IQ	4.102***	10.95	0.000		-		-
	Slider	3.688***	25.93	0.000	3.09	0.079		-
	Grade	3.089***	58.27	0.000	18.60	0.000	6.52	0.011
					Men			
	Random	4.564***		-		-		-
(3)	IQ	4.439***	0.09	0.767		-		-
	Slider	3.808***	3.44	0.064	2.35	0.126		-
	Grade	3.605***	5.32	0.211	4.12	0.042	0.26	0.613
					Women			
	Random	5.018***		-		-		-
(4)	IQ	4.000***	13.06	0.000		-		-
	Slider	3.596***	25.48	0.000	2.03	0.154		-
	Grade	2.807***	61.79	0.000	17.87	0.000	7.88	0.005

Table notes: Rating scale from 1 (= completely not legitimate) to 7 (= completely legitimate) as dependent variable. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Explanation: The third row displays the margins of each mechanism for the three distinct tobit regressions (Model 2 to 4). The last 6 rows includes mutual tests of significant differences between mechanisms. In other words, a Wald-F-test is estimated to test the null hypothesis that two mechanisms are rated equally/similar. For example, in the second row of the first part of the table, the rating of IQ is compared to random and tested if the evaluation between IQ and random equals zero. For this comparison, the $\chi^2 = 10.95$ and the $p - value = 0.000$, so that the null hypothesis has to be rejected. Consequently, the rating of Random and IQ differ significantly.

A.3. Part 1 - Dictator Game with Punishment

Table A.2.: Simple Dictator Game with Punishment

Punishment Points	(1)	(2)	(3)	(4)
Default	-0.285*	-0.264*		
	(0.160)	(0.157)		
Mechanism	-0.179	-0.154		
	(0.160)	(0.157)		
Grade-Unequal			-0.169	-0.139
			(0.223)	(0.219)
Random-Equal			-0.275	-0.248
			(0.223)	(0.219)
Grade-Equal			-0.464**	-0.418*
			(0.223)	(0.219)
Allocation	2.885***	2.885***	2.885***	2.885***
	(0.149)	(0.149)	(0.149)	(0.149)
Choice as Dictator		0.479***		0.479***
		(0.160)		(0.160)
Female		-0.197		-0.198
		(0.165)		(0.165)
Age		0.013		0.013
		(0.015)		(0.015)
Constant	-2.344***	-2.870***	-2.349***	-2.877***
	(0.262)	(0.497)	(0.272)	(0.502)
Wald-χ^2	377.41***	388.82***	377.41***	388.82***
$p(\chi^2)$	0.000	0.000	0.000	0.000
N	540	540	540	540

Table notes. Random-effects tobit regression. Dependent variable: Punishment points (from 0 to 5). Mechanism equals 1 if grade is the group-building mechanism. Default equals 1 if the preselected option is equal. Random-unequal is the baseline category. Allocation equals 1 if the punishment for the unequal allocation is assigned. Choice as dictator equals one if the equal allocation is chosen as dictator. Female is a dummy variable for the respective socio-demographic variable. Standard errors in parentheses $^{*}p \leq 0.1$, $^{**}p \leq 0.05$,$^{***}p \leq 0.01$.

A.4. Further Regression Analyses

Table A.3.: Correlationsmatrix

	Choice default	Female	Age	Economists	Belief	Deserve Role A	Deserve Role B	Rating Random	Rating Grade	Risk attitude	Random more	Grade more
Choice default	1											
Female	0.0176	1										
	0.2196											
Age	-0.0311*	-0.0808*	1									
	0.0303	0.0000										
Economists	0.0079	-0.1143*	-0.0649*	1								
	0.5800	0.0000	0.0000									
Belief	-0.0279*	-0.0076	0.0368*	-0.0142	1							
	0.0519	0.5959	0.0102	0.3235								
Deserve Role A	0.0017	-0.0003	0.0196	-0.0445*	-0.0147	1						
	0.9042	0.9832	0.1723	0.0019	0.3051							
Deserve Role B	0.0342*	0.0695*	0.0425*	-0.0011	0.0460*	0.5788*	1					
	0.0172	0.0000	0.0030	0.9387	0.0013	0.0000						
Rating Random	0.0205	0.2067*	-0.0473*	-0.0712*	0.0150	0.0650*	0.1240*	1				
	0.1536	0.0000	0.0010	0.0000	0.2973	0.0000	0.0000					
Rating Grade	-0.0316*	-0.1135*	-0.0114	-0.0152	0.0121	0.2774*	0.1730*	-0.1432*	1			
	0.0275	0.0000	0.4257	0.2904	0.3974	0.0000	0.0000	0.0000				
Risk attitude	-0.0165	-0.1936*	0.0543*	0.1119*	-0.0300*	0.1033*	-0.0025	-0.1337*	0.0162	1		
	0.2515	0.0000	0.0002	0.0000	0.0366	0.0000	0.8639	0.0000	0.2603			
Random more	0.0451*	0.1994*	-0.1591*	-0.0484*	-0.0226	-0.0160	-0.0017	0.6123*	-0.5947*	-0.1304*	1	
	0.0016	0.0000	0.0000	0.0007	0.1152	0.2639	0.9053	0.0000	0.0000	0.0000		
Grade more	-0.0667*	-0.1831*	-0.0328*	0.0671*	-0.0079	0.0491*	-0.1053*	-0.4350*	0.5220*	0.1666*	-0.5739*	1
	0.0000	0.0000	0.0224	0.0000	0.5813	0.0006	0.0000	0.0000	0.0000	0.0000	0.0000	
Outcome unequal	-0.0186	-0.0395*	-0.0133	0.0408*	-0.0701*	-0.0398*	-0.0717*	-0.0118	0.0319*	0.0121	-0.0352*	0.0356*
	0.1956	0.0059	0.3524	0.0044	0.0000	0.0055	0.0000	0.4117	0.0263	0.3998	0.0142	0.0131
Choice unequal	-0.0242*	-0.0162	-0.0154	0.0369*	-0.0996*	-0.0285*	-0.0494*	-0.0086	0.0196	0.0088	-0.0213	0.0201
	0.0919	0.2586	0.2830	0.0101	0.0000	0.0471	0.0000	0.5475	0.1724	0.5388	0.1371	0.1602
Unkind	-0.0231	-0.0153	-0.0113	0.0334*	-0.0863*	-0.0284*	-0.0481*	-0.0127	0.0145	0.0090	-0.0215	0.0182
	0.1076	0.2848	0.4322	0.0199	0.0000	0.0479	0.0000	0.3771	0.3110	0.5325	0.1344	0.2057
Initiator	0.0162	-0.0008	-0.0110	0.0245*	-0.0690*	-0.0161	-0.0291*	-0.0084	0.0035	0.0046	-0.0080	0.0059
	0.2579	0.9528	0.4415	0.0878	0.0000	0.2614	0.0424	0.5567	0.8055	0.7471	0.5762	0.6794

Table continues

	Choice default	Female	Age	Economists	Belief	Deserve Role A	Deserve Role B	Rating Random	Rating Grade	Risk attitude	Random more	Grade more
Pivotal	-0.0596*	-0.0211	-0.0071	0.0218	-0.0598*	-0.0213	-0.0384*	-0.0063	0.0170	0.0065	-0.0188	0.0190
	0.0000	0.1409	0.6191	0.1282	0.0001	0.1379	0.0075	0.6608	0.2349	0.6527	0.1898	0.1851
Outcome X Choice	-0.0880*	-0.0336*	-0.0123	0.0348*	-0.0664*	-0.0329*	-0.0569*	-0.0063	0.0320*	0.0104	-0.0308*	0.0316*
	0.0000	0.0191	0.3925	0.0154	0.0000	0.0220	0.0001	0.6627	0.0259	0.4666	0.0320	0.0275
Outcome X Initiator	-0.0596*	-0.0211	-0.0071	0.0218	-0.0248*	-0.0213	-0.0384*	-0.0063	0.0170	0.0065	-0.0188	0.0190
	0.0000	0.1409	0.6191	0.1282	0.0834	0.1379	0.0075	0.6608	0.2349	0.6527	0.1898	0.1851
BSJO equity	0.0618*	0.0309*	-0.1130*	0.0937*	-0.0235	0.2882*	0.2915*	-0.0574*	0.2372*	0.0683*	-0.0896*	0.1004*
	0.0000	0.0311	0.0000	0.0000	0.1016	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
BSJO equality	0.0125	0.0424*	0.1108*	-0.1233*	0.0208	-0.1556*	-0.0892*	0.0286*	-0.2170*	-0.0644*	0.0890*	-0.1775*
	0.3837	0.0031	0.0000	0.0000	0.1473	0.0000	0.0000	0.0465	0.0000	0.0000	0.0000	0.0000
BSJO need	0.0071	0.0224	0.0042	-0.2078*	0.0179	0.0171	-0.0577*	0.1488*	-0.1487*	-0.0142	0.1734*	-0.1375*
	0.6183	0.1190	0.7692	0.0000	0.2111	0.2326	0.0001	0.0000	0.0000	0.3210	0.0000	0.0000
BSJO entitlement	0.0248*	-0.2111*	-0.0253	0.0687*	-0.0097	0.2218*	0.2044*	-0.1679*	0.3549*	0.1388*	-0.2097*	0.2254*
	0.0840	0.0000	0.0777	0.0000	0.4974	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table continues

	Outcome unequal	Choice unequal	Unkind	Initiator	Pivotal	Outcome X Choice	Outcome X Initiator	BSJO equity	BSJO equality	BSJO need	BSJO entitlement
Choice default											
Female											
Age											
Economists											
Offer B											
Deserve Role A											
Deserve Role B											
Rating Random											
Rating Grade											
Random vs. grade											
Random more											
Grade more											
Outcome unequal	1										
Choice unequal	0.5850*	1									
	0.0000										
Unkind	0.5981*	0.9340*	1								
	0.0000	0.0000									
Initiator	0.2714*	0.7788*	0.7179*	1							
	0.0000	0.0000	0.000								

Table continues

	Outcome unequal	Choice unequal	Unkind	Initiator	Pivotal	Outcome X Choice	Outcome X Initiator	BSJO equity	BSJO equality	BSJO need	BSJO entitlement
Pivotal	0.5345*	0.5091*	0.5451*	-0.1093*	1						
	0.0000	0.0000	0.0000	0.0000							
Outcome X Choice	0.8053*	0.7670*	0.7761*	0.3847*	0.6638*	1					
	0.0000	0.0000	0.0000	0.0000	0.0000						
Outcome X Initiator	0.5345*	0.5091*	0.5091*	0.6537*	-0.0714*	0.6638*	1				
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000					
BSJO equity	-0.0246*	-0.0196	-0.0171	-0.0137	-0.0132	-0.0190	-0.0132	1			
	0.0863	0.1711	0.2327	0.3407	0.3592	0.1860	0.3592				
BSJO equality	-0.0171	-0.0110	-0.0082	-0.0052	-0.0091	-0.0144	-0.0091	-0.3548*	1		
	0.2340	0.4414	0.5689	0.7178	0.5247	0.3139	0.5247	0.0000			
BSJO need	-0.0610*	-0.0364*	-0.0416*	-0.0177	-0.0326*	-0.0488*	-0.0326*	0.0244*	0.2509*	1	
	0.0000	0.0112	0.0037	0.2171	0.0230	0.0007	0.0230	0.0886	0.0000		
BSJO entitlement	0.0312*	0.0202	0.0223	0.0066	0.0167	0.0293*	0.0167	0.2486*	-0.4175*	-0.4094*	1
	0.0294	0.1592	0.1201	0.6448	0.2444	0.0409	0.2444	0.0000	0.0000		

Pairwise correlation, *p < 0.1

OLS Regression

Table A.5.: OLS Regression with Control Variables

Punishment Point	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mechanism	0.008						
	(0.099)						
Default		0.201**					
		(0.081)					
Grade-unequal			-0.018	-0.025	0.060	0.066	0.108
			(0.064)	(0.076)	(0.062)	(0.119)	(0.102)
Random-equal			0.179	0.320	0.406	0.214	0.216
			(0.159)	(0.414)	(0.420)	(0.241)	(0.230)
Grade-equal			0.206***	0.337	0.353	0.108	0.136
			(0.062)	(0.337)	(0.330)	(0.156)	(0.183)
Choice default				-0.246	-0.248	-0.050	-0.049
				(0.588)	(0.578)	(0.163)	(0.163)
Female					-0.239*	-0.212	-0.258*
					(0.130)	(0.142)	(0.148)
Age					0.012	0.017	0.020
					(0.017)	(0.015)	(0.015)
Risk affine					-0.092	-0.008	-0.004
					(0.139)	(0.129)	(0.107)
Belief					-0.245***	-0.033	-0.033
					(0.048)	(0.032)	(0.032)
Economist					-0.186*	-0.225**	-0.213**
					(0.106)	(0.109)	(0.084)
Deserve role A					-0.001	0.004	-0.013
					(0.026)	(0.026)	(0.024)
Deserve role B					-0.032	-0.018	-0.015
					(0.028)	(0.026)	(0.024)
Rating random					-0.054*	-0.044	-0.046
					(0.028)	(0.029)	(0.028)
Rating grade					-0.009	0.014	0.011
					(0.021)	(0.018)	(0.020)
Outcome unequal						-0.079	-0.075
						(0.061)	(0.062)
Choice unequal						1.533***	1.532***
						(0.225)	(0.226)
Outcome X Initiator						0.749***	0.750***
						(0.256)	(0.257)
Pivotal						0.265	0.267
						(0.301)	(0.302)
BSJO equity							0.052
							(0.044)
BSJO equality							-0.083
							(0.069)
BSJO need							0.186**
							(0.086)
BSJO entitlement							0.008
							(0.052)
Constant	0.513***	0.419***	0.427***	0.476***	0.852*	0.082	-0.655
	(0.086)	(0.033)	(0.061)	(0.127)	(0.476)	(0.435)	(0.473)
r2-overall	0.000	0.006	0.006	0.013	0.045	0.330	0.341
N	2430	2430	2430	2430	2430	2430	2430

Table notes: Dependent variable: Punishment points (from 0 to 7). Random effects OLS regression, clustered on sessions and with standard errors in parentheses. Mechanism equals 1 if grade is the group-building mechanism. Default equals 1 if the preselected option is equal. Random-unequal is the baseline category. Choice default equals 1 if the preselected option is chosen. The variables 'Female', 'Economics', and 'RIsk affine' are dummy variables for the respective socio-demographic variable. Deserve Role A and B range from 1 (completely not deserving) to 10 (completely deserving). Rating random and grade range from 1 (completely not legitimate) to 7 (completely legitimate). Punishment motives are dummy variables for the respective category as defined in section 7.4.1. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$..

Tobit Regression

Table A.6.: Testing for a Mechanism Effect - Marginal Effects

	Initiation		Pivotality	
	No	Yes	No	Yes
Random not more legitimate	0.310**	1.991***	0.405***	0.707***
	(0.155)	(0.235)	(0.155)	(0.141)
Random more legitimate	0.516***	1.009***	0.529***	0.809***
	(0.067)	(0.140)	(0.067)	(0.141)

Table notes: Average marginal effect of tobit regressions (Model 13 and 14) with standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$.

Table A.7.: Tobit Regression - Robustness checks

Punishment Points	Only round 1	Without round 1	Without round 6	Punishment Points 0-6	Men	Female	Only unequal outcome
Grade-unequal	0.187	0.092	0.151	0.093	0.459*	-0.016	0.096
	(0.199)	(0.145)	(0.147)	(0.142)	(0.249)	(0.153)	(0.373)
Random-equal	0.363*	0.209	0.329**	0.246*	0.593**	0.100	0.185
	(0.207)	(0.139)	(0.141)	(0.135)	(0.252)	(0.144)	(0.330)
Grade-equal	0.049	0.166	0.125	0.148	0.348	-0.060	0.171
	(0.208)	(0.146)	(0.147)	(0.142)	(0.239)	(0.158)	(0.338)
Choice default	-0.104	-0.070	-0.098	-0.078	-0.220**	0.048	-0.076
	(0.141)	(0.056)	(0.061)	(0.049)	(0.086)	(0.063)	(0.140)
Female	-0.228	-0.266**	-0.211*	-0.248**	/	/	-0.397
	(0.154)	(0.112)	(0.113)	(0.109)	/	/	(0.248)
Age	0.011	0.021**	0.018*	0.017*	0.038**	0.013	0.016
	(0.012)	(0.009)	(0.009)	(0.009)	(0.018)	(0.009)	(0.020)
Risk affine	0.060	-0.022	0.029	-0.015	-0.048	0.121	0.274
	(0.167)	(0.121)	(0.123)	(0.119)	(0.190)	(0.147)	(0.307)
Belief	0.047	-0.042	-0.033	-0.021	-0.168*	0.038	-0.229
	(0.150)	(0.056)	(0.066)	(0.050)	(0.089)	(0.065)	(0.217)
Economist	0.023	-0.258**	-0.163	-0.197	-0.244	-0.113	-0.261
	(0.173)	(0.125)	(0.127)	(0.123)	(0.186)	(0.146)	(0.284)
Deserve role A	-0.009	-0.014	-0.018	-0.007	-0.049	0.029	-0.056
	(0.036)	(0.026)	(0.026)	(0.025)	(0.046)	(0.029)	(0.068)
Deserve role B	-0.020	-0.014	-0.016	-0.020	-0.008	-0.037	0.030
	(0.032)	(0.023)	(0.023)	(0.023)	(0.036)	(0.027)	(0.063)
Rating random	-0.093**	-0.043	-0.069**	-0.039	-0.008	-0.074**	-0.208***
	(0.042)	(0.029)	(0.030)	(0.029)	(0.047)	(0.034)	(0.067)
Rating grade	0.029	0.005	0.007	0.007	0.021	0.029	0.032
	(0.045)	(0.033)	(0.033)	(0.032)	(0.052)	(0.038)	(0.084)
Outcome unequal	-0.628***	-0.042	-0.112	-0.079	-0.068	-0.076	/
	(0.240)	(0.090)	(0.102)	(0.078)	(0.136)	(0.104)	/
Choice unequal	2.101***	1.430***	1.617***	1.526***	1.554***	1.520***	1.185***
	(0.195)	(0.080)	(0.087)	(0.069)	(0.126)	(0.090)	(0.402)
Outcome X Initiator	0.857***	0.729***	0.750***	0.474***	0.903***	0.629***	1.059***
	(0.311)	(0.139)	(0.153)	(0.121)	(0.212)	(0.159)	(0.392)
Pivotal	-0.177	0.356**	0.294*	0.203*	0.166	0.305*	0.561
	(0.311)	(0.139)	(0.152)	(0.120)	(0.211)	(0.158)	(0.391)
BSJO equity	0.021	0.060	0.049	0.041	-0.109	0.121*	0.248*
	(0.088)	(0.064)	(0.064)	(0.062)	(0.131)	(0.065)	(0.139)
BSJO equality	-0.059	-0.087	-0.062	-0.079	-0.240**	-0.060	-0.002
	(0.079)	(0.057)	(0.057)	(0.056)	(0.100)	(0.061)	(0.143)
BSJO need	0.322***	0.167**	0.197**	0.174**	0.528***	0.005	0.157
	(0.118)	(0.085)	(0.086)	(0.083)	(0.137)	(0.100)	(0.182)
BSJO entitlement	0.020	0.005	0.038	0.021	0.256**	-0.243***	-0.235
	(0.100)	(0.072)	(0.073)	(0.071)	(0.104)	(0.094)	(0.175)
Constant	-0.834	-0.580	-0.656	-0.562	-2.257**	0.209	-0.099
	(0.820)	(0.594)	(0.602)	(0.582)	(0.933)	(0.673)	(1.455)
Wald-χ^2	262.60	1112.50	1008.01	1319.00	585.48	848.45	267.81
$p(\chi^2)$	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AIC	1269.12	5771.36	4705.38	6678.18	2593.02	4381.26	1795.22
N	405	2025	1620	2419	882	1548	486

Table notes: Dependent variable: Punishment points (from 0 to 7). Random effects tobit regression with standard errors in parentheses. Mechanism equals 1 if grade is the group-building mechanism. Default equals 1 if the preselected option is equal. Random-unequal is the baseline category. Choice default equals 1 if the preselected option is chosen. The variables 'Female', 'Economics', and 'RIsk affine' are dummy variables for the respective socio-demographic variable. Deserve Role A and B range from 1 (completely not deserving) to 10 (completely deserving). Rating random and grade range from 1 (completely not legitimate) to 7 (completely legitimate). Punishment motives are dummy variables for the respective category as defined in section 7.4.1. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$..

BSJO-scale

Table A.8.: Operationalization of the BSJO-scale

Justice principle	Item	Item text
Equity	C	It is just if all people have the same living conditions
(Gleichheit)		(Gerecht ist, wenn alle die gleichen Lebensbedingungen haben)
	K	It is just if income and wealth are equally distributed among the members of our society
		(Es ist gerecht, wenn Einkommen und Vermögen in unserer Gesellschaft an alle Personen gleich verteilt werden)
Need	E	A society is just if it takes care of those who are poor and needy
(Bedarf)		(Eine Gesellschaft ist gerecht, wenn sie sich um die Schwachen und Hilfsbedürftigen kümmert)
	J	It is just if people taking care of their children or their dependent relatives receive special support and benefits
		(Es ist gerecht, wenn Personen, die Kinder oder pflegebedürftige Angehörige zu versorgen haben, besondere Unterstützung und Vergünstigungen erhalten)
Equity	B	It is just if hard working people earn more than others
(Leistung)		(Es ist gerecht, wenn Personen, die hart arbeiten, mehr verdienen als andere)
	I	It is just if every person receives only that which has been acquired through their own efforts
		(Gerecht ist, wenn jede Person nur das bekommt, was sie sich durch eigene Anstrengungen erarbeitet hat)
Entitlement	D	It is just if members of respectable families have certain advantages in their lives
(Anrecht)		(Es ist gerecht, wenn Personen, die aus angesehenen Familien stammen, dadurch Vorteile im Leben haben)
	L	It is fair if people on a higher level of society have better living conditions than those on the lower level
		(Es ist gerecht, wenn diejenigen, die in einer Gesellschaft oben stehen, bessere Lebensbedingungen haben als diejenigen, die unten stehen)

Notes: Items were asked in alphabetical order. German translation, as used in this study in parentheses. The introductory text reads: "There are different ideas about how a society can be fair and just. What is your personal opinion about this?" / "Es gibt unterschiedliche Vorstellungen darüber, wann eine Gesellschaft gerecht ist. Wie ist Ihre persönliche Meinung dazu? Bitte geben Sie an, ob Sie jeweils: voll zustimmen, etwas zustimmen, weder zustimmen noch ablehnen, etwas ablehnen oder ganz ablehnen"

Source: Based on Hülle et al. (2018)

Table A.9.: Descriptive Statistics of Items and Dimensions of the BSJO-scale in LINOS-1 and this Study

		LINOS-1		This study	
	Item	Mean	SD	Mean	SD
Equality	C	3.30	1.30	3.67	1.29
	K	2.26	1.26	2.69	1.31
	Subscale	2.78	1.09	3.18	1.11
Need	E	4.56	0.64	4.41	0.89
	J	4.60	0.63	4.48	0.82
	Subscale	4.58	0.50	4.44	0.73
Equity	B	4.48	0.74	4.01	1.04
	I	3.57	1.17	3.03	1.20
	Subscale	4.03	0.76	3.52	0.94
Entitlement	D	1.62	0.92	1.75	1.06
	L	2.29	1.16	2.02	1.11
	Subscale	1.95	0.85	1.89	0.95

Table notes: SD = standard deviation. Scale ranging from (1)“strongly disagre” to (5) “strongly agre”.
LINOS-1: N = 4509; this study: N = 270.

Table A.10.: Top-5 Answers of Recipients by Default, Mechanism, and Treatments

			Default		Mechanism	
	Answer	Overall	Unequal	Equal	Random	Grade
1.	punishment costs / is expensive	28.24%	23.88%	32.81%	23.88%	32.81%
2.	unequal choice	23.66%	22.39%	25.00%	17.91%	29.69%
3.	fairness	21.37%	26.87%	15.63%	25.37%	17.19%
4.	equal choice	19.85%	17.91%	21.88%	20.90%	18.75%
5.	punishment brings no advantage	17.56%	10.45%	25.00%	14.93%	20.31%

Table notes: Multiple answers are possible.

Comparison to Bartling et al. (2015)

Table A.11.: Comparison to Bartling et al. (2015) - Random Treatments

Allocation	Voting sequence	Voter 1	Voter 2	Voter 3	Pivotal vs. all	Pivotal vs. unkind	Unkind vs. all
		Bartling et al. (2015)			**Wilkoxon sign rank**		
Unequal	u-u-u	1.50	1.85	0.86	p<0.001	p=0.294	
	u-u-e	1.86	1.92	0.26	p<0.001	p=0.960	
	u-e-u	1.68	0.07	2.39	p<0.001	p=0.006	
	e-u-u	0.11	1.83	2.33	p<0.001	p=0.012	
Equal	u-e-e	1.33	0.10	0.08			
	e-u-e	0.17	1.43	0.08			
	e-e-u	0.06	0.03	0.92			
	e-e-e	0.08	0.07	0.03			
		This study			**Wilkoxon ranksum**		
Unequal	u-u-u	1.11	0.89	1.00	p=0.977	p=0.920	p=0.885
	u-u-e	2.42	2.25	0.17	p=0.033	p=0.617	p=0.006
	u-e-u	2.59	0.19	2.26	p=0.021	p=0.654	p=0.003
	e-u-u	0.00	2.67	1.60	p=0.271	p=0.154	p=0.004
Equal	u-e-e	1.36	0.14	0.19			
	e-u-e	0.10	1.31	0.10			
	e-e-u	0.05	0.48	2.57			
	e-e-e	0.17	0.12	0.16			

Notes: "u" denotes a choice of the unequal allocation; "e" denotes a choice of the equal allocation. The three rightmost columns show p-values of Wilcoxon signed rank tests for Bartling et al. (2015) and a ranksum test for this study comparing the punishment for the pivotal decision-maker to the punishment for the two other decision-makers ("pivotal vs. all") and to the punishment for the other intentionally unkind decision-maker only ("pivotal vs. unkind"). Additionally, in this study a comparison between the punishment for the unkind decision-maker to the punishment for the two other decision-makers ("unkind vs. all") is added.

Table A.12.: Comparison of Different Punishment Motives - Tobit Regression

Punishment Points	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Outcome unequal	1.176***								-0.030
	(0.064)								(0.087)
Unequal choice		1.804***							1.097***
		(0.050)							(0.267)
Unkind			1.755***						
			(0.054)						
Equal choice				-1.804***					
				(0.050)					
Initiator					1.673***				0.467*
					(0.063)				(0.278)
Outcome X Initiator						1.857***			0.692***
						(0.094)			(0.133)
Outcome X Choice							1.802***		
							(0.066)		
Pivotal								1.312***	0.677**
								(0.098)	(0.265)
Constant	0.282***	0.127**	0.177***	1.931***	0.277***	0.393***	0.265***	0.429***	0.128**
	(0.052)	(0.052)	(0.052)	(0.064)	(0.051)	(0.051)	(0.052)	(0.051)	(0.052)
Sigma u	0.521***	0.543***	0.542***	0.543***	0.533***	0.525***	0.539***	0.525***	0.542***
	(0.040)	(0.039)	(0.039)	(0.039)	(0.040)	(0.040)	(0.040)	(0.041)	(0.039)
Sigma e	1.131***	0.970***	1.003***	0.970***	1.058***	1.119***	1.052***	1.166***	0.959***
	(0.017)	(0.014)	(0.015)	(0.014)	(0.016)	(0.017)	(0.016)	(0.017)	(0.014)
N	2430	2430	2430	2430	2430	2430	2430	2430	2430

Table notes: Random effects tobit regression with punishment points (from 0 to 7) as dependent variable. All punishment motives are dummy variables. Outcome unequal equals 1 if the unequal allocation is chosen. Unkind equals 1 if a decision-maker opts for the unequal allocation. Initiation equals 1 if the decision-maker is the first one choosing unequal. Pivotal equals 1 if a decision-maker is pivotal for the unequal outcome. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A.13.: Comparison of different Punishment Motives - Random Treatments

Punishment	This study									Bartling et al.
Points	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(2015)
Outcome unequal	1.286***								0.031	0.048
	(0.109)								(0.096)	(0.070)
Unequal choice		1.790***							1.228***	0.782***
		(0.251)							(0.212)	(0.137)
Unkind			1.688***							
			(0.175)							
Equal choice				-1.790***						
				(0.251)						
Initiator					1.587***				0.201	0.517**
					(0.315)				(0.350)	(0.196)
Outcome X Initiator						1.921***			0.851***	0.289
						(0.309)			(0.314)	(0.232)
Outcome X Choice							1.891***			
							(0.193)			
Pivotal								1.459***	0.639***	0.403**
								(0.212)	(0.207)	(0.155)
Constant	0.280***	0.132***	0.195***	1.922***	0.283***	0.397***	0.271***	0.425***	0.127***	0.083**
	(0.065)	(0.037)	(0.054)	(0.276)	(0.045)	(0.068)	(0.058)	(0.078)	(0.042)	(0.037)
R^2-overall	0.137	0.312	0.258	0.312	0.201	0.135	0.224	0.082	0.328	0.281
N	1242	1242	1242	1242	1242	1242	1242	1242	1242	1728

Notes: Random effects OLS regression with punishment points (from 0 to 7) for decision-makers as dependent variable. All punishment variables are dummy variables. Outcome unequal equals 1 if the unequal allocation is chosen. Unkind equals 1 if a decision-maker opts for the unequal allocation. Initiation equals 1 if the decision-maker is the first one choosing unequal. Pivotal equals 1 if a decision-maker is pivotal for the unequal outcome. Standard errors in parentheses, clustered on 12 session in this study and 72 individuals in Bartling et al. (2015); ***p<0.01, **p<0.05, *p<0.1.

Expected Punishment

Table A.14.: Tobit Regression - Expected Punishment

Expected Punishment	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mechanism	-0.077						
	(0.110)						
Default		0.269**					
		(0.108)					
Grade-unequal			-0.257*	-0.266*	-0.259*	-0.159	-0.150
			(0.149)	(0.150)	(0.145)	(0.100)	(0.100)
Random-equal			0.096	0.278*	0.304**	0.136	0.128
			(0.149)	(0.155)	(0.150)	(0.103)	(0.103)
Grade-equal			0.196	0.365**	0.348**	0.139	0.120
			(0.149)	(0.155)	(0.156)	(0.108)	(0.108)
Choice default				-0.318***	-0.315***	-0.122***	-0.120***
				(0.072)	(0.072)	(0.045)	(0.045)
Female					0.082	0.029	0.012
					(0.113)	(0.078)	(0.080)
Age					0.003	-0.002	-0.005
					(0.010)	(0.007)	(0.007)
Risk affine					0.201*	0.064	0.074
					(0.111)	(0.077)	(0.077)
Economist					0.199	0.092	0.104
					(0.123)	(0.085)	(0.088)
Deserve role A					-0.013	-0.009	-0.008
					(0.022)	(0.015)	(0.015)
Deserve role B					-0.004	0.006	0.011
					(0.025)	(0.017)	(0.019)
Rating random					0.010	-0.009	-0.008
					(0.030)	(0.021)	(0.022)
Rating grade					0.086***	0.019	0.030
					(0.031)	(0.021)	(0.023)
Outcome unequal						-0.031	-0.028
						(0.073)	(0.073)
Choice unequal						3.249***	3.250***
						(0.064)	(0.064)
Outcome X Initiator						-0.569***	-0.569***
						(0.111)	(0.111)
Pivotal						-0.982***	-0.982***
						(0.111)	(0.111)
BSJO equity							-0.016
							(0.042)
BSJO equality							0.046
							(0.035)
BSJO need							-0.014
							(0.054)
BSJO entitlement							-0.012
							(0.044)
Constant	0.808***	0.639***	0.762***	0.826***	0.378	0.225	0.225
	(0.077)	(0.075)	(0.103)	(0.105)	(0.378)	(0.260)	(0.393)
Wald-χ^2	0.49	6.23	9.80	28.96	/	/	/
$p(\chi^2)$	0.485	0.013	0.020	0.000	/	/	/
AIC	8682.50	8676.89	8677.52	8660.33	8661.99	6285.41	6290.39
N	2430	2430	2430	2430	2430	2430	2430

Table notes: Dependent variable: expected punishment points (from 0 to 7). Random effects tobit regression with standard errors in parentheses. Mechanism equals 1 if grade is the group-building mechanism. Default equals 1 if the preselected option is equal. Random-unequal is the baseline category. Choice default equals 1 if the preselected option is chosen. The variables 'Female', 'Economics', and 'Risk affine' are dummy variables for the respective socio-demographic variable. Deserve Role A and B range from 1 (completely not deserving) to 10 (completely deserving). Rating random and grade range from 1 (completely not legitimate) to 7 (completely legitimate). Punishment motives are dummy variables for the respective category as defined in section 7.4.1. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$..