Strobel, Christina; Kirchkamp, Oliver

Conference Paper

Sharing responsibility with a machine

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Experiments - Games II, No. B18-V1

Provided in Cooperation with:
Verein für Socialpolitik / German Economic Association

Sharing Responsibility with a Machine*

Oliver Kirchkamp,† Christina Strobel‡

August 1, 2017

Humans take decisions jointly with others. They share responsibility for the outcome with their interaction partners. More and more often the partner in this decision is not another human but, instead, a machine. Here we ask whether a machine partner affects our responsibility, our perception of the choice and our choice itself differently than a human partner. As a workhorse we use a modified dictator game with two joint decision makers: either two humans or one human and one machine.

We find a strong treatment effect on perceived responsibility. We do, however, find only a small and insignificant effect on actual choices.

Keywords: Experiment, Dictator Game, Human-Machine Interaction, Hybrid-Decision Situation

*This document has been generated on August 1, 2017, with R version 3.4.0 (2017-04-21), on x86_64-w64-mingw32. We would like to thank the Max Planck Society for financial support through the International Max Planck Research School on Adapting Behavior in a Fundamentally Uncertain World. We would also like to thank the audience of the IMPRS internal doctoral seminars for their input. Data and methods are available at `https://www.kirchkamp.de/research/shareMachine.html`.

†FSU Jena, School of Economics, Carl-Zeiss-Str. 3, 07737 Jena, `oliver@kirchkamp.de`.

‡FSU Jena, School of Economics, Bachstraße 18k, 07737 Jena, `Christina.Strobel@uni-jena.de`.

1. Introduction

In almost all areas of life decisions are more and more the result of an interaction of humans with a machine. We find automated systems no longer only in a supportive function but, more frequently, as systems which take actions on their own: Computer assisted driving services drive autonomously on roads, surgical systems conduct surgeries independently, etc.[1] [2] As a result humans find themselves confronted with a new situation: they have to share decisions with a machine. We call such a situation a hybrid decision situation.

This paper aims to investigate human decision making in hybrid decision situations. More specifically, we want to investigate if sharing a decision with a computer instead of with another human changes the perception of the situation and the actual decision. Findings from economics (see Engel, 2011; Luhan et al., 2009) and social psychology (see Darley and Latané, 1968; Wildschut et al., 2003) suggest that humans decide in a more selfish way if a decision is shared with another human. We ask whether we find a similar pattern if a decision is shared with a machine. We also investigate if the perceived responsibility (see Fischer et al., 2011; Latané and Nida, 1981) and the perceived guilt (see Battigalli and Dufwenberg, 2007; Rothenhäusler et al., 2013) is affected by the type of interaction partner.

As a workhorse we use a binary dictator game and compare three treatments: a dictator game with a single dictator, a dictator game with two human dictators, and a dictator game with one human and one machine dictator.

The remainder of the paper is organized as follows: Section 2 provides a literature review focusing on experimental evidence from economics and social psychological research. We present studies on individual behavior in groups as well as findings from research on human-computer interactions. Section 3 presents the experimental design. Section 4 relates the experiment to the theoretical background and derives behavioral predictions. Results are presented in Section 5.The last section offers a discussion and some concluding remarks.

2. Review of the Literature

Human decision making in groups with other humans has been researched extensively in economics as well as in social psychology. In Section 2.1 we focus on the question why humans behave more selfishly when deciding with other humans. In Section 2.2 we summarise research on human-computer interactions.

2.1. Shared Decision Making with Humans

People often have to make decisions in situations where the overall outcome does not only depend on their own decisions but also on the decisions of others. A stable experimental

[1] See for example Choi et al. (2016); Seaman (2016); Senthilingam (2016); Stone et al. (2016).

[2] In several situations humans are outperformed by machines. These are situations where machines are more accurate or reliable. Also, machines do not seem to show signs of boredom. Nevertheless, machines not not necessarily perform better than humans. For example, according to an international survey 56.8% of 176 responding surgeons had experienced an irrecoverable intraoperative malfunction of the robotic system during an urological surgery (Kaushik et al., 2010).

observation is that humans are more selfish, less altruistic and are less trustworthy if they decide together with others.[3] Already in a very simple game, the Dictator Game, people seem to behave in a more strategic and selfish way when deciding in groups compared to individual decision making. Dana et al. (2007) find that in a situation where two dictators decide simultaneously and where the selfish outcome is implemented only if both dictators agree on it, 65% of all dictators choose the selfish option, while only 26% of all dictators choose the selfish option when deciding alone. This observation is confirmed by Luhan et al. (2009) where 23.4% of a person's endowment is sent to a responder when people decide alone but only 19% is sent when people act as members of a three-person team. [4] Experiments in social psychology also find that people are less likely to help when others are around and are more likely to give less money when part of a group (Panchanathan et al., 2013).[5]

Although experimental evidence shows that people behave more self-seeking in shared decisions not much research has been done to investigate the driving forces behind it. According to Falk and Szech (2013) and Bartling et al. (2015) individuals behave in a more selfish way when deciding in groups as the pivotality for the final outcome is diffused. This diffusion facilitates to opt for a self interested option as the individual perception of being decisive for the final outcome is lowered. Battigalli and Dufwenberg (2007) provide another explanation by arguing that human actions are influenced by the aim to reduce the feeling of guilt caused by a decision. Rothenhäusler et al. (2013) transfer this idea to group decisions by stating that the possibility to share the guilt for a decision with others facilitates to act selfishly.

There are also psychological concepts which offer explanations for more selfish behavior in groups. Darley and Latané (1968) propose the concept of *diffusion of responsibility*: selfish decisions in groups are caused by the possibility to share the responsibility for the outcome among group members. This idea is confirmed by several studies which show that people indeed tend to feel less responsible for the final outcome when they have to decide together with others (see Darley and Latané, 1968; Forsyth et al., 2002; Freeman et al., 1975; Latané and Nida, 1981; Wallach et al., 1964). Research on the so called *interindividual-intergroup discontinuity effect*, which describes the tendency of individuals to be more competitive and less cooperative in groups than in one-on-one relations, has obtained further possible mechanisms which could be driving more selfish decision-making in groups (see Wildschut et al., 2003). First, the social-support-for-shared-self-interest hypothesis states that group members can perceive an active support for a self-interested choice by other group members. Second, the identifiability hypothesis proposes that deciding in groups provides a shield of anonymity which could also drive selfish decision-making. Third, the ingroup-favouring norm could put some pressure on decision makers to behave in a way which benefits the group before taking

[3]This has been shown in a number of experimental games such as the Trust Game (Kugler et al., 2007), the Ultimatum Game (Bornstein and Yaniv, 1998), the Coordination Game (Bland and Nikiforakis, 2015), the Signaling Game (Cooper and Kagel, 2005), the Prisoners Dilemma (McGlynn et al., 2009), the Gift Exchange Game (Kocher and Sutter, 2007), the Public Good Games (Andreoni and Petrie, 2004) as well as in lotteries (Rockenbach et al., 2007) and Beauty Contests (Kocher and Sutter, 2005; Sutter, 2005).

[4]Bland and Nikiforakis (2015) study a coordination game with third party externalities and find selfish behaviour among the joint decision makers even when the selfish option imposes a strong negative externality on a third-party.

[5]For an overview on the so called *bystander-effect* see Fischer et al. (2011).

into account the interests of others. Finally, the altruistic-rationalization hypothesis states that group members can justify their own selfish behavior by arguing that the other dictator is also benefiting from the decision.

2.2. Perception of and Behavior towards Computers

While research in economics and social psychology analyzes shared decision making between humans, there seems to be a gap when it comes to individuals who share decisions with a machine instead of with another human.

A number of studies find that machines are treated in a way similar to humans. Nass and Moon (2000) find that humans ascribe human-like attributes to machines. Humans also apply social rules and expectations to machines.[6] Moon and Nass (1998) observe that humans have a tendency to blame a computer for failure and take the credit for success when they feel dissimilar to it while blaming themselves for failure and crediting the computer for success when they feel similar to it. In addition, several studies find that computers are held at least partly responsible for actions (see Bechel, 1985; Friedman, 1995; Moon, 2003).

Although humans sometimes seem to treat computers and humans in a similar way, differences remain: Melo et al. (2016) find that behavior towards humans differs from behavior towards machines. Humans contribute more money to a public good when the good is shared with humans than with machines. Humans also offer more money to human responders in an Ultimatum Game than to an artificial counterpart. Humans also expect more money from machines than from humans in a modified Dictator Game. Melo et al. also find that people are more likely to perceive guilt when interacting with an human counterpart than when interacting with machines. Envy, however, does not seem to depend on the type of the opponent.

Especially in domains in which fundamental human properties such as moral considerations and ethical norms are of importance, findings from human-human interactions can not necessarily directly transferred to human-computer interactions. Gogoll and Uhl (2016) find that people seem to dislike the usage of computers in moral domains where a decision also affects another person. In their experiment people were able to delegate their decision in a trust game either to a human or to a computer algorithm which exactly resembles the human behavior in a previous trust game. The fact that only 26.52% of all subjects delegated their decision to the computer while 73.48% delegated their decision to a human shows that people are reluctant to delegate to a machine. Gogoll and Uhl also allowed impartial observers to reward or to punish actors conditional on their delegation decision. They find that, independent of the outcome, impartial observers reward delegations to a human more than delegation to a computer.

[6] Humans also seem to respond socially to computers (Katagiri et al., 2001; Reeves and Nass, 2003), use social rules in addressing computer behavior (Nass et al., 1994), apply human stereotypes to machines (Eyssel and Hegel, 2012) and accept computers as teammates (Nass et al., 1996).

3. Experimental Design

We implement an experimental design with the following elements: (i) a binary Dictator Game in which people are able to choose between an equal and an unequal split, (ii) a questionnaire to measure the preceived responsibility and guilt, and (iii) a manipulation check in which people were confronted with a counterfactual deciding situation, The decision in the Dictator Game is made either by a single human dictator (SDT), by two (multiple) human dictators (MDT), or by a computer together with a human dictator (CDT).

3.1. General Procedures

In each experimental session, the following procedure was used: Upon arrival at the laboratory participants were randomly seated and randomly assigned a role (Player X, Player Y, and, in existing in the treatment, Player Z). All participants were informed that they would be playing a game with one or two other participants in the room and that matching would be randomly and anonymously. They were also told that all members of all groups would be paid according to the choices made in that group. Payoffs were explained using a generic payoff table. A short quiz ensured that the task and the payoff representation was understood. After passing the quiz the actual payoffs for the experiment were shown to participants together with any other relevant information for the treatment.

All treatments were one-shot dictator games with a binary choice between an equal and an unequal (welfare inefficient) wealth allocation. After making the choice and before being informed about the final outcome subjects were asked to answer a questionnaire to determine their perceived level of responsibility and guilt. Every participant was paid privately on exiting the room. All experimental stimuli as well as instructions were presented via a computer interface. We framed the game as neutral as possible, avoiding any loaded terms.

The entire experiment was computerized using z-Tree (Fischbacher, 2007). All subjects were recruited via ORSEE (Greiner, 2004).

3.2. Treatments

A a between subject design was used to compare three different treatments: One treatment involves two players, a single dictator and a single responder (Single Dictator Treatment, SDT). Two more treatments involve three players, two dictators and one responder. While in one of these all players are human (Multiple Dictator Treatment, MDT), one of the dictators is replaced by a computer in the second treatment (Computer Dictator Treatment, CDT).

3.2.1. Single Dictator Treatment

Payoffs for the Single Dictator Treatment (SDT) are shown in the left part of Table 1. The dictator (Player X) can choose either an unequal allocation (Option A) with a higher gain for her or an equal allocation (Option B). In this treatment we call Player Y the responder.

Single Dictator Treatment:

Player X's choices		
	A	Y:1 X:6
	B	Y:5 X:5

Multiple Dictator and Computer Dictator Treatments:

Player X's choices		Player Y's choices: **A**		**B**	
	A	Y:6 X:6	Z:1	Y:5 X:5	Z:5
	B	Y:5 X:5	Z:5	Y:5 X:5	Z:5

Table 1: Binary Dictator Game

3.2.2. Multiple Dictator

Payoffs for the Multiple Dictator Treatment (MDT) are shown in the right part of Table 1. Dictators (Player X and Player Y) both make a choice which determines the payoff of the dictator and the responder (Player Z). The unequal payoff is only implemented if both dictators choose Option A. In all other cases, Option B is implemented.

3.2.3. Computer Dictator

The Computer Dictator Treatment (CDT) is identical to the MDT with one exception: Player Y acts as a so called passive dictator with his choice being made by a computer, called in the following computer dictator. The computer dictator chooses Option A with the fraction of dictators who had chosen Option A in an earlier MDT. Participants in the CDT are told that the probability with which the computer decides between Option A and Option B follows the behavior of participants in a former experiment. Hence, all Player X in the CDT treatment have the same beliefs (and the same ambiguity) about the other player. To be comparable with the MDT, Player Y in the group was paid as in the MDT, also this Player Y had no influence on the allocation.

3.3. Measurement of Perceived Responsibility and Guilt

In each treatment participants had to answer a questionnaire before being informed about the payoff. We elicited the perceived responsibility for the outcome as well as feeling of guilt caused by a selfish decision.

After choosing an option but before being informed about the final outcome and payoff, dictators in each treatment were asked to state their perceived responsibility for the outcome in a case where Option A would be finally implemented as well as their perceived responsibility for payoff of the responder and for the payoff of the dictator(s). Dictators were also asked to state how guilty they would feel if Option A would be implemented. These questions were used as a proxy for the perceived responsibility for the final outcome as well as for the level of perceived guilt caused by a self-interested decision. Similar to Luhan et al. (2009) all participants were also asked to state their reasons for choosing a specific option.

Furthermore, in MDT and CDT dictators and responders were asked to state the expected behavior of the other players as well as how they allocate the responsibility and guilt for

the decision between the dictators or correspondingly the computer and the dictator. [7] In addition, dictators had the opportunity to state the reason why they chose a specific option in an open question and correspondingly the responders had the opportunity to state why they expected the dictator(s) to choose a specific option.

We also conducted a manipulation check by asking questions about how participants would evaluate the situation used in another treatment. Furthermore, we collected some demographic data. Data and methods are available online.[8]

4. Theoretical Framework and Behavioral Hypotheses

From a traditional economic point of view humans should aim to maximize their monetary utility without taking the welfare of others into account or being influenced by situational circumstances. Thus, their decision should be based on maximizing their own profit, regardless of whether they have to decide on their own, with another person or with a computer. However, unlike traditional economic theories predict, the decision situation seems to influence individual behavior. As experimental studies have shown subjects behave more self-oriented and less pro-social the less salient the link between ones actions and the consequences for someone else is. This holds especially true in situations where the final outcome is influenced by ambiguity or uncertainty (see Chen and Schonger, 2013; Haisley and Weber, 2010) or when more or less plausible excuses to justify a self-interested behavior are available (see Grossman and van der Weele, 2013; Grossman, 2014; Matthey and Regner, 2011). Models of social image concerns (see Andreoni and Bernheim, 2009; Bénabou and Tirole, 2006; Ellingsen and Johannesson, 2008; Grossman, 2015) and models on self-perception maintenance (see Aronson, 2009; Beauvois and Joule, 1996; Bodner and Prelec, 2003; Konow, 2000; Mazar et al., 2008; Murnighan et al., 2001; Rabin, 1995) provide a theoretical basis for this finding. According to these models, individuals do not only want to maximize their own output but also want to be perceived by others as kind and fair as well as being able to see themselves in a positive light. However, if these two goals are at odds, opting for an option which maximizes one's own output causes an unpleasant tension for the individual which can only be reduced by lowering the perceived conflict of interest between the two goals.[9] Thereby, as research in social psychology has shown, people seem to act selectively and in a self-serving way when determining whether a self-interested behavior will have a positive or negative impact on their own self-concept or social image and use situational excuses, if available, to justify their decision (see Rabin, 1995; Haidt and Kesebir, 2010). This allows to attribute selfish actions to the context, instead of having to attribute it to the own self-concept and thus facilitates to opt for a selfish option as it enables to uphold a comfortable self- and social image.

Applied to a situation where the decision is shared this means that with an increasing number of deciders involved in the decision, the perceived personal responsibility a single

[7] Responders were asked for their expectations of the responsibility and guilt felt by the dictators.

[8] `https://www.kirchkamp.de/research/shareMachine.html`

[9] The unpleasant tension (or in a more formal speech "disutility") is often described as nothing else than the feeling of guilt (see Berndsen and Manstead, 2007; de Hooge et al., 2011; Stice, 1992).

dictator might feel for the final outcome decreases. Due to this the dictator might only feel responsible for a fraction of the harm caused by his self-interested decision. Furthermore, as stated by Berndsen and Manstead (2007) and Bruun and Teroni (2011), the perceived feeling of guilt for a selfish decision will also be reduced if the perceived personal responsibility for the final outcome is lowered. In addition, the diffused pivotality due to the uncertainty of the own decision being finally determinant also provides an excuse to feel less responsible for the final outcome (see Bartling et al., 2015; Falk and Szech, 2013). Thus, sharing a decision with another human makes it easier to choose a more self-serving option as it allows to reduce the perceived negative consequences for the self- and social image.

As in our experiment choosing Option B leads to an equal output for all participants it causes less harm to the social and self-image than choosing Option A, where the responder receives much less than the dictator(s). Thus, dictators who value a positive perception by others and themselves higher [lower] than maximizing their monetary output will always choose Option B [A]. However, dictators who strive to maximize their own profits while at the same time maintaining a positive self- and social image, are facing two conflicting desires: maximize their own output by choosing Option A or maintain a positive self- and social image by choosing Option B.

In the SDT the payoff depends only on the choice of the dictator and thus offers no situational excuse to reduce the negative impact on the self- and social image caused by a selfish decision. Sharing a decision with another dictator (as in the MDT) provides some moral wiggle room to interpret a selfish behavior as more favorably for the majority and thus allows to attribute a selfish decision to the context or situational circumstance instead of to one's own self- and social image.[10] Based on that we expect that dictators in the MDT perceive themselves to be less responsible for the final outcome (Hypothesis 1.i) as well as to feel less guilty for a selfish decision (Hypothesis 2.i) than dictators in the SDT. As a result we expect more selfish decisions in the MDT than in the SDT (Hypothesis 3.i).

When looking at the CDT it becomes clear that it makes no sense to held the computer dictator as responsible as a human dictator. Research has come up with multiple conditions that need to be fulfilled to be held responsible for an action. The following three conditions can be seen as the main ones even though researchers vary in the degree to which they have to be fulfilled. First, an agent needs to have action power, meaning there has to be a causal relationship between his actions and the outcome (see Lipinski et al., 2002; May, 1992; Moore, 1999; Nissenbaum, 1994; Scheines, 2002). Second, it is required to be able to choose freely, including the competence to act on the basis of own authentic thoughts and motivations [11] as well as the capability to control one's own behavior [12]. Third, to be held responsible requires the ability to consider the possible consequences an action might cause (see Bechel, 1985; Friedman and Kahn, 1992). Some researchers even argue that it is necessary to be capable to suffer or gain from possible blame or praise and thus being culpable for wrongdoing (see Moor et al., 1985; Sherman, 1999; Wallace, 1994). These conditions would also have to hold true for a computer to be held responsible. As the causal responsibility of a computer for

[10]However, as either dictator can independently implemented the equal outcome by choosing Option B the addition of a second dictator does not impede subjects from ensuring a fair outcome if they prefer it.

[11]For the so called *freedom of will* condition see Fischer, 1999

[12]For the so called *freedom of action* condition see Johnson, 2006

an outcome cannot be denied it neither has a free will nor the freedom of action (see Floridi and Sanders, 2004; Johnson and Powers, 2005; Sparrow, 2007) and is also not able to consider possible consequences of its actions in the same way as a human (see Bechel, 1985; Friedman and Kahn, 1992; Moon and Nass, 1998). A computer is also not capable of any kind of own emotions (see Asaro, 2011; Snapper, 1985; Sparrow, 2007). This illustrates that a computer does not fulfill the conditions under which it would make sense to hold it responsible to the same extent as a human.[13]

Based on this the responsibility for a selfish outcome can not be shared with a computer to the same extent as with a human and the wiggle room is smaller than in a shared decision with another human. Thus, upholding a positive self- and social image while deciding selfishly together with a computer should not be as easy as when deciding with another human. For this reasons we expect dictators to perceive a higher level of own responsibility for the final outcome (Hypothesis 1.ii) and guilt when choosing the unfair option (Hypothesis 2.ii) in the CDT than in the MDT. In addition, as selfish decision making is influenced by the individual's perception of being responsible or feeling guilty for a decision, significantly more people should choose the selfish option if they are deciding with another human (MDT) compared to with a computer (CDT) (Hypothesis 3.ii).

Hypothesis 1 *In a situation where the outcome depends on the decision of two humans (Multiple Dictator Treatment) participants do allocate less responsibility for the outcome resulting from choosing the selfish option than*

(i) if the outcome is determined by a single dictator (Single Dictator Treatment) or, alternatively,

(ii) if the outcome depends on the decision of a human and a computer (Computer Dictator Treatment).

Hypothesis 2 *In a situation where the outcome depends on the decision of two humans (Multiple Dictator Treatment) participants do allocate less guilt for the outcome resulting from choosing the selfish option than*

(i) if the outcome is determined by a single dictator (Single Dictator Treatment) or, alternatively,

(ii) if the outcome depends on the decision of a human and a computer (Computer Dictator Treatment).

Hypothesis 3 *In a situation where the outcome depends on the decision of two humans (Multiple Dictator Treatment) the selfish option is chosen more often than*

(i) if the outcome is determined by a single dictator (Single Dictator Treatment) or, alternatively,

[13] This is also supported by research in machine and robot ethics which only attributes operational responsibility to the most advanced machines today but denies any higher form of (moral) responsibility as today's machines are still having a relatively low level of own autonomy and ethical sensitivity (see Allen et al., 2000; DeBaets, 2014; Dennett, 1997; Moor et al., 1985; Sullins, 2006).

	multiple/single	multiple/computer
dictator	[-Inf,-11.37] (0.0000)	[-Inf,5.914] (0.3495)
responder	[-Inf,1.855] (0.1042)	[-Inf,0.6293] (0.0662)

Table 2: Treatment difference in the responsibility of the human dictator(s) as seen by dictators and responders

(ii) if the outcome depends on the decision of a human and a computer (Computer Dictator Treatment).

5. Results

All sessions were run in July, October and November 2016 at the Friedrich Schiller Universität Jena. Three treatments were conducted with a total of 399 subjects (65.2% female).[14] Most of our subjects were students with an average age of 25 years. Participants earned in the experiment on average €9.43. The data for all statistical tests is independent for the different treatments as we applied a between-subject design. We first analyze how the perceived responsibility for the final outcome as well as the feeling of guilt for a self-serving decision varied between the treatments before presenting the findings regarding the choices made by the dictators.

5.1. Responsibility

Different kinds of responsibilities need to be considered to check how the perceived responsibility for a selfish decision differs between the treatments. All subjects were asked to state the level of responsibility for the final outcome that they impose on each of the dictators (human in SDT, humans in the MDT, and human and computer in the CDT).[15] To evaluate if the perceived responsibility towards the responder (and if present the other dictator) differs between the treatments, subjects were also asked to evaluate how responsible they feel for the final payoff of the responder (and if existing the other dictator) in each treatment.[16] For all questions the level of responsibility was measured by a continuous scale from *"Not responsible at all"* (0) to *"Very responsible"* (100).

[14] 124 subjects (62.9% female) participated in the SDT, 92 subjects (68.5% female) in the MDT and 183 subjects (65% female) in the CDT. We have, thus, almost the same number of active dictators in each treatment (see Table 5).

[15] For the exact wording see Question 9 from Section A.1.2.

[16] Dictators were asked how responsible they feel, responders and passive dictators were asked how responsible they perceive the dictator to be. For the exact wording see Question 6 and Question 7 from Section A.1.2.

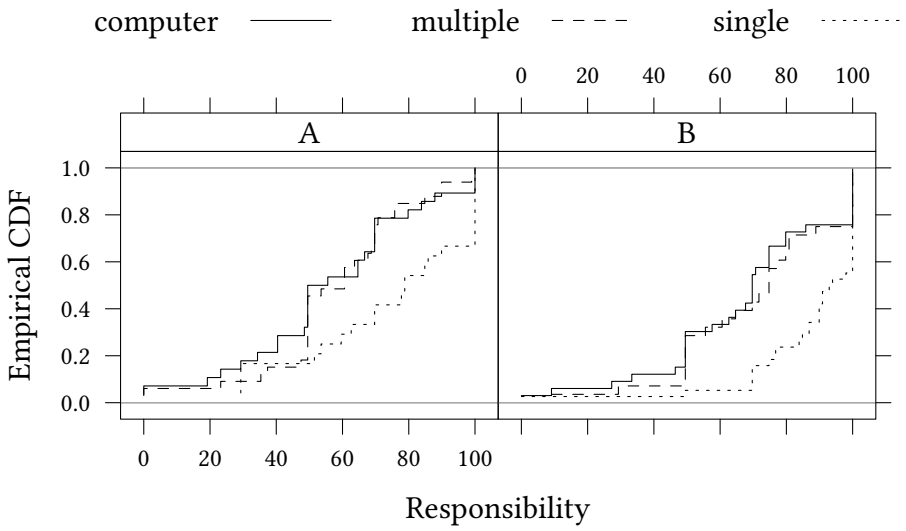

Figure 1: Perceived own responsibility as seen by dictators
(Question 9 from Section A.1.2)

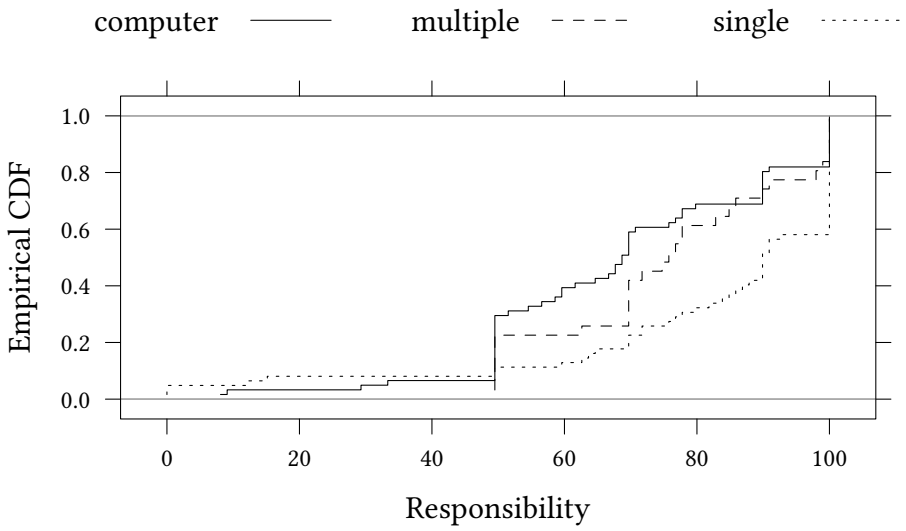

Figure 2: Responsibility allocated to the human dictator as seen by responders
(Question 9 from Section A.1.2)

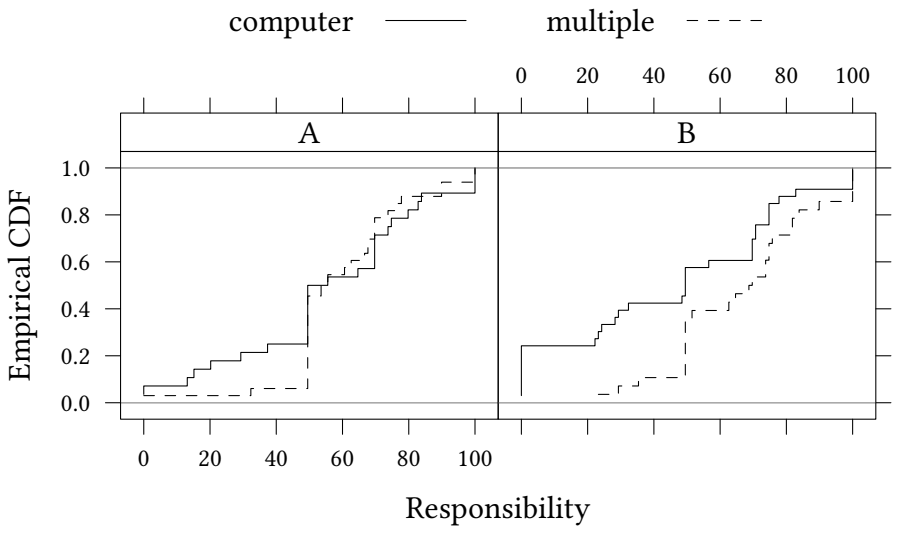

Figure 3: Responsibility allocated to the other eighter human or computer dictator as seen by dictators
(Question 9 from Section A.1.2)

5.1.1. Responsibility of (First) Dictator

Figure 1 shows the perceived own responsibility for the final outcome by the dictator(s).[17] In Figure 2 the responsibility allocated by the responders to the human dictator(s) is shown. [18] As Table 2 [19] shows we can clearly confirm Hypothesis 1.i for the dictators, i.e. that the responsibility for the outcome perceived by the decision maker is lower if the decision is shared with another human compared to when deciding alone. However, the responsibility allocated to the dictator(s) by the responders did not differ significantly between the SDT and the MDT. Thus, we can not confirm Hypothesis 1.i for the responders.

The same figures and tables can be used to analyse1.ii. While the preceived responsibility for the final decision did not differ significantly for dictators, the responders perceived the human dictator in the MDT as slightly more responsible for the decision than the computer dictator in the CDT. Thus, Hypothesis 1.ii, i.e. that the own responsibility for the outcome as perceived by the deicsion maker is lower if the decision is shared with another human than with a computer, can not be confirmed for dictators but weakly be confirmed for responders.

5.1.2. Responsibility of the Other Dictator

The other dictator was eighter a human (in the MDT) or a computer (in the CDT). Figure 3 as well as Figure 4 show a significant difference in responsibility allocated to the other dictator by dictators who chose Option B, and also by responders between the MDT and

[17] The Figure is split up, sharing the results for dictators who have chosen Option A on the left and dictators who have chosen Option B on the right.
[18] Passive dictator responses are analyzed in the Appendix.
[19] The following p-values are based on t-tests, unless stated otherwise.

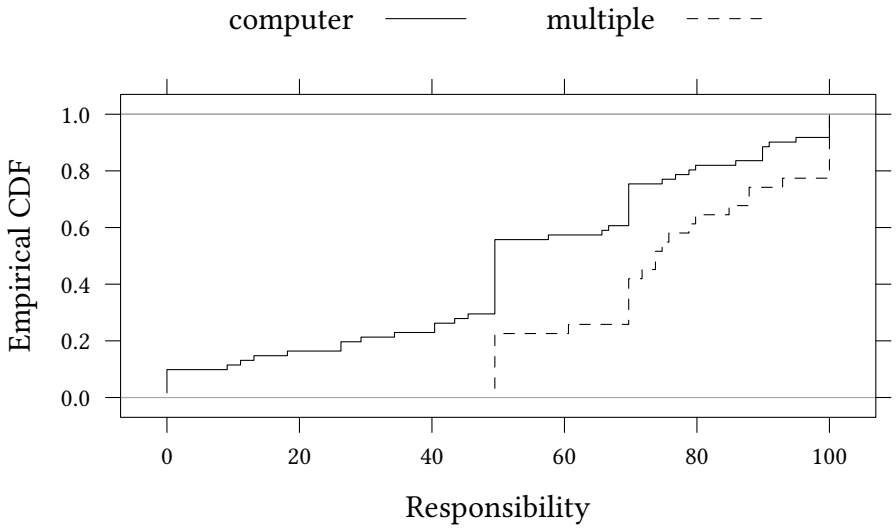

Figure 4: Responsibility allocated to the other eighter human or computer dictator as seen by responders
(Question 9 from Section A.1.2)

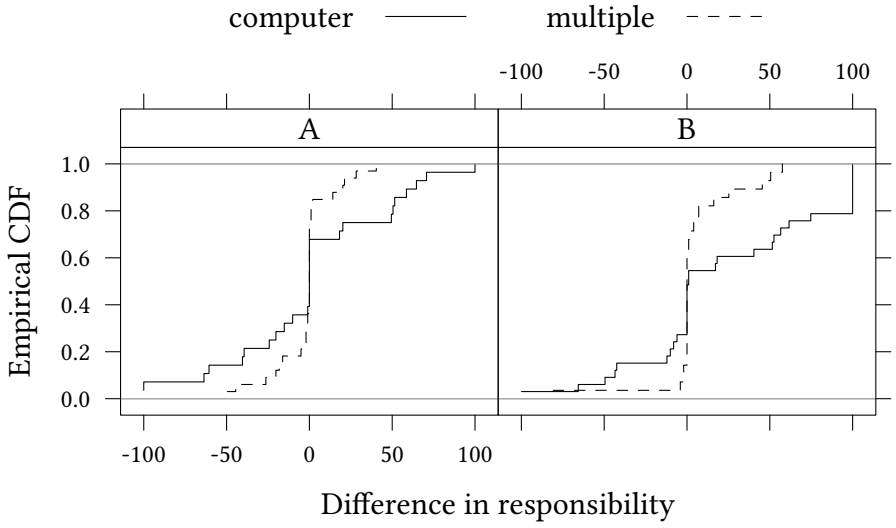

Figure 5: Difference in responsibility of the dictator between themselves and the other eighter human or computer dictator as seen by the dictators
(Question 9 from Section A.1.2)

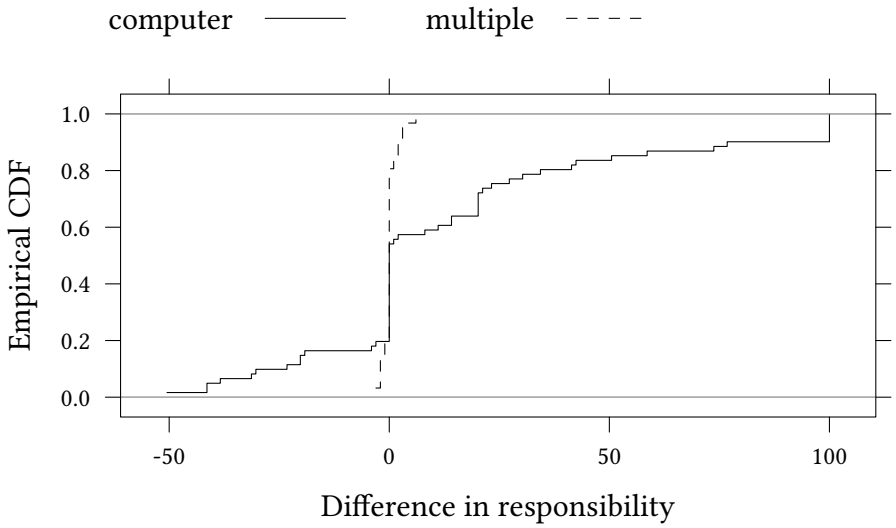

Figure 6: Difference in responsibility of the human dictator and the other eighter human or computer dictator as seen by the responders
(Question 9 from Section A.1.2)

	multiple/computer
dictator	[-Inf,-4.935] (0.0042)
responder	[-Inf,-13.06] (0.0000)

Table 3: Treatment difference between the responsibility of the human dictator and the other eighter human or computer dictators' responsibility as seen by dictators and responders

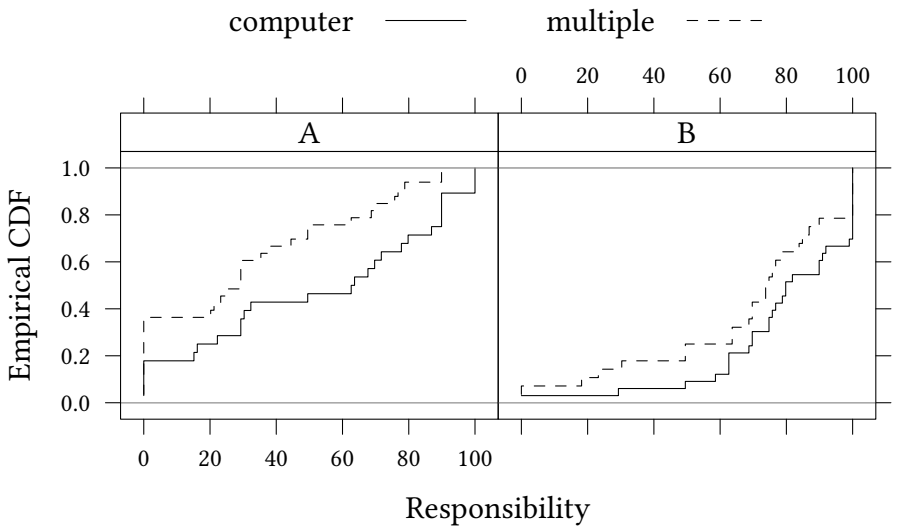

Figure 7: Own responsibility for the other eighter active or passive dictator as seen by dictators
(Question 7 from Section A.1.2)

	multiple/single resp.	multiple/computer resp.	multiple/computer pass.
dictator	[-Inf,2.473] (0.1269)	[-Inf,5.747] (0.2814)	[8.547,Inf] (0.0014)
responder	[-2.956,Inf] (0.2614)	[-Inf,-8.782] (0.0000)	[-Inf,6.779] (0.3505)

Table 4: Treatment difference in responsibility for responders and passive dictators as seen by dictators and responders

the CDT. Interestingly, dictators who chose Option B as well as responders perceived the human dictator in the MDT on average as significantly more responsible for the final outcome than the computer in the CDT as Table 3 shows. [20] When comparing the responsibility the dictators allocate to themselves with the responsibility the dictators allocate to the the other decider (see Figure 5) it becomes clear that the difference is more dispersed in the CDT, where dictators had to decide together with a computerized dictator, than in the MDT, where dictators decided together with another human dictator. [21] The same holds true for responders as Figure 6 shows. [22]

5.1.3. Responsibility for Others

Dictators dictators were asked to state how responsible they feel for the final payoff of the responder and the payoff of the other dictator, if present. In addition we asked the respon-

[20] For dictators this effect is maily driven by dictators who chose Option B.
[21] Means, however, are similar (p-value 0.0637).
[22] However, the difference between the responsibility allocated by the responder to the first and second dictator is clearly and significantly more dispersed in the CDT (p-value 0.0017).

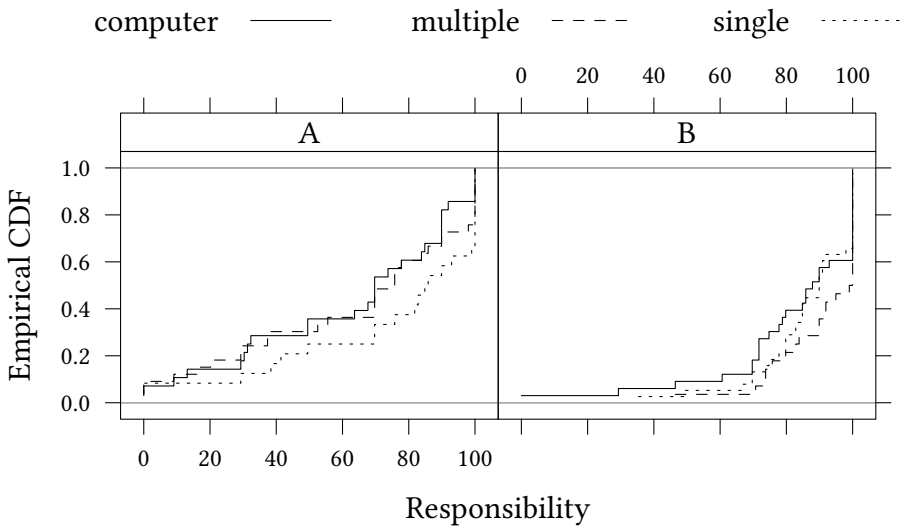

Figure 8: Own responsibility for the responder as seen by dictators
(Question 6 from Section A.1.2)

ders to evaluate the level of responsibility they think the dictators perceive for the payoff of the responder and, if present, the other dictator.[23] The own responsibility perceived by the dictators for the other dictator and for the responder are shown in Figures 7 and Figure 8.[24] As shown in Table 4 dictators stated to perceive a significantly higher level of responsibility for the payoff of the passive dictator in the CDT than for the other actively deciding dictator in the MDT. However, their perceived level of responsibility for the responders' payoff did not differ significantly between the treatments. [25] Interestingly, responders expected rather the opposite. They estimated that dictators would perceive themselves to be significantly more responsible for the responders payoff in the CDT than in the MDT but did not expect the same for the perceived responsibility of the dictator for the payoff of the other dictator.

5.1.4. Responsibility Findings

The findings regarding the responsibility can be summed up in three points. (a) Dictators perceived themselves on average as significantly less responsible for their decision in the MDT than in the SDT but the perceived level of own responsibility for the decision did not differ significantly between the CDT and the MDT. Responders, however, expected that dictators perceive themselves to be more responsible in the CDT than in the MDT (weak significant), but did not show a significant difference for the passive dictator between the perceived responsibility in the MDT compared to the SDT. Furthermore, (b) dictators as well as responders allocated less responsibility to the computer in the CDT than to the other human

[23] For the exact wording of the question see Question 6 from Section A.1.2.
[24] For the corresponding figures of the responders evaluation see Figure 24 and Figure 25 in Section A.4.3.
[25] It is important to note that the overall level of responsibility perceived by dictators for the responders' payoff is higher for dictators who chose Option B compared to dictators who chose Option A, in all treatments.

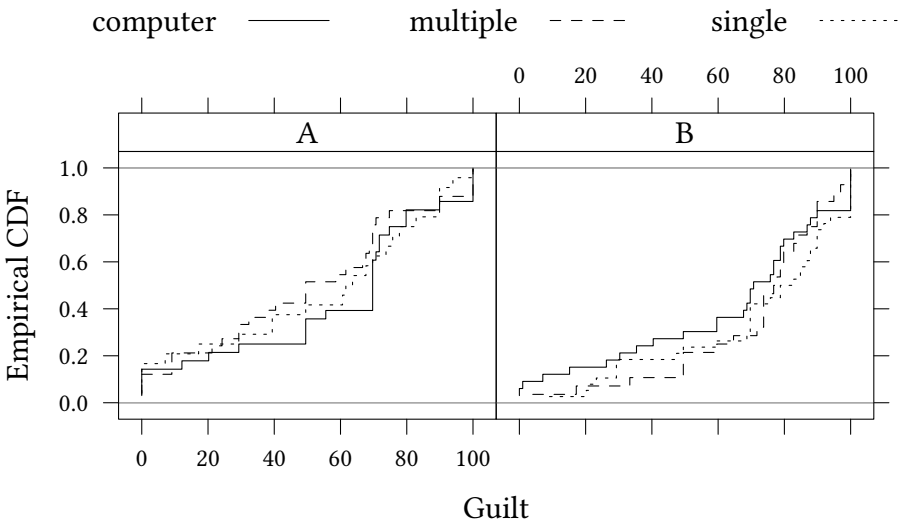

Figure 9: Perceived guilt as seen by dictators
(Question 8 from Section A.1.2)

dictator in the MDT. Finally, (c) dictators felt significantly more responsible for the payoff of the passive dictator in the CDT than for the other actively deciding dictator in the MDT. This was not expected by the responders as they expected that the dictators would perceive themselves to be significantly more responsible for the responders payoff in the CDT than in the MDT but did not show a significant difference in their expectation towards the dictators perception of responsibility. Thus, Hypothesis 1.i, i.e. that the perceived own responsibility for the outcome by the decision maker is lower if the decision is shared with another human can be confirmed for dictators but not for responders. Hypothesis 1.ii, i.e. that the perceived own responsibility by the decision maker for the outcome is lower if the decision is shared with another human than with a computer can not be confirmed for dictators but weakly be confirmed for responders. Interestingly, dictators as well as responders perceived aa computer to be less responsible than a human dictator. Furthermore, while responders expected dictators to perceive more responsibility for the payoff of the responderin the CDT than in the MDT, dictators stated to feel significantly more responsible for the payoff of the passive dictator in the CDT than in the MDT.

5.2. Guilt

In all treatments dictators were asked to state their perceived guilt in case option A was going to be implemented.[26] In addition, responders were asked to state how guilty they expect the dictator(s) to feel for the final payoff. The level of guilt was measured by a continuous scale from *"not guilty"* (0) to *"totally guilty"* (100). According to Hypothesis 2.i, we expect the perceived guilt for the outcome to be lower in the MDT than in the SDT. Furthermore, as

[26]For the exact wording of the question see Question 8 from Section A.1.2.

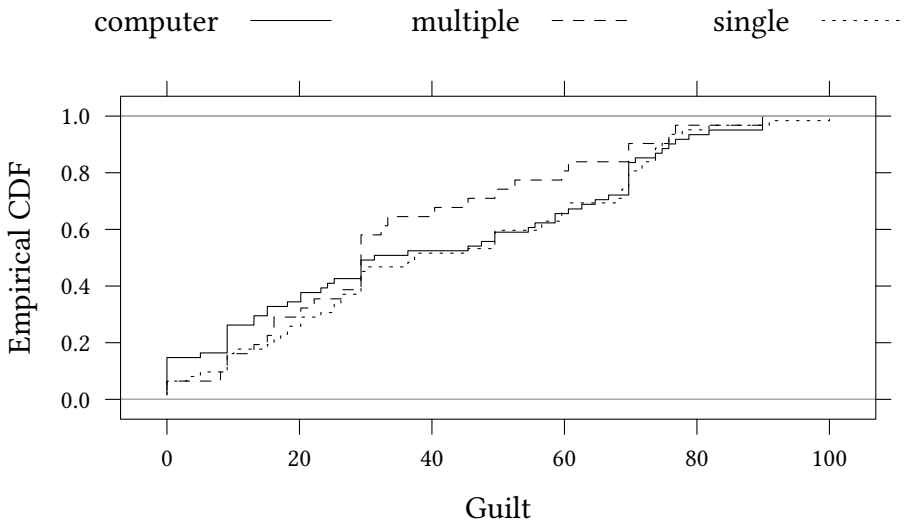

Figure 10: Expected guilt of the dictator(s) as seen by responders
(Question 8 from Section A.1.2)

stated in Hypothesis 2.ii, we expect a lower level of perceived guilt if the outcome depends on the decision of two humans, as in the MDT, than if the outcome depends on the decision of a human and a computer, as in the CDT.

The dictators' perceived level of guilt, if Option A would be implemented, is shown in Figure 9. Apparently, Hypothesis 2.i, that dictators feel less guilty if the outcome depends on the decision of two humans compared to an outcome determined by a single dictator, can not be confirmed (p-value 0.1875). The same also applies for Hypothesis 2.ii which states that dictators feel less guilty when choosing the selfish option in the MDT compared to the CDT (p-value 0.4344).

Figure 10 shows that the guilt allocated by the responders to the dictator(s) for the final payoff did not differ significantly between the treatments (p-value 0.7664). Thus, neither Hypothesis 2.i nor Hypothesis 2.ii can be confirmed for dictators nor responders..

To sum up, dictators did not feel significantly more or less guilty for a selfish decision if they had to decide on their own, together with a computer or another human. Responders also expected no significant difference in the dictators' perceived level of guilt for a selfish decision between the treatments.

5.3. Choices and Expected Choices

The number of selfish choices made by the dictators varies among the different treatments. [27] An overview of the selfish-choices per treatment can be found in Table 5. We find only weak support for Hypothesis 3.i, i.e. that when the outcome depends on the decision of two humans the selfish option is chosen more often than if the outcome is determined by a single

[27]For the binary Dictator Game interface shown to the dictators and to the responders see Section A.1.1.

Treatment	Proportion choosing A
computer dictator treatment	28/61 (45.9%)
multiple dictator treatment	33/61 (54.1%)
single dictator treatment	24/62 (38.7%)

Table 5: Number of selfish choices by treatments
(for the Question see Figure 11 in Section A.1)

exp. no. of A choices	computer	multiple	single
0	37.7	6.5	64.5
1	62.3	29.0	35.5
2	0.0	64.5	0.0

Table 6: Responders' expectations of "A" choices [%]
(for the Question see Figure 12 in Section A.1)
Note that in the single and computer treatments there is only a single opponent, hence, there can be no more than one A choice.

dictator. (p-value[28] 0.0630). Regarding Hypothesis 3.ii, i.e. that when the outcome depends on the decision of two humans the selfish option is chosen more often than if the outcome depends on the the decision of a human and a computer, we see that the proportion of selfish choices in CDT is somewhere between SDT and MDT. This is in line with our hypothesis. The effect is, however, small and not significant (p-value for more selfishness in MDT than in CDT is 0.2344, for more selfishness in CDT than in SDT is 0.2661).

Table 6 summarises the responders' expectations for the number of Option A-choices in the three treatments. Indeed, responders expected significantly more selfish choices (per dictator) in MDT than in SDT (p-value[29] 0.0001). Furthermore, expectations in the CDT were between expectations for MDT and SDT. Responders expect fewer selfish choices (per dictator) in CDT than in MDT (p-value 0.0544). Responders expected even fewer selfish choices in the SDT (p-value 0.0017).

6. Conclusion

The number of decisions made by human-computer teams have already increased substantially in the past and will continue to increase in the future. Here we study whether humans perceive a decision shared with a computer differently from a decision shared with another human. More specifically, we focus on responsibility and guilt. From other studies we know that humans behave more selfishly if they share responsibility with other humans. We can replicate this finding in our experiment, even for human-computer interactions. We find that, if responsibility is shared with a machine, more selfish choices are made than if decisions are taken alone but fewer than if decisions are taken together with a human. Differences in

[28]The p values in this paragraph are based on tests for proportionality.
[29]The p-values in this paragraph are based on a logistic model.

actual choices are not significant. We do, however, get significant effects in expectations: Essentially, sharing responsibility with a machine does not permit as much selfish behaviour as sharing it with a human, but definitely much more than when deciding alone.

We investigate two potential reasons why humans may expect fewer selfish decisions if the decision is shared with a machine: responsibility and guilt. Own responsibility is clearly perceived as smaller once a second decision maker comes into play. It does not matter whether the second decision maker is a machine or a human. Although our participants attribute more responsibility to a human counterpart, their own responsibility is reduced both with a computer and a human partner by very similar amounts. Guilt does not seem to be affected by the type of the interaction.

Our results underline the importance of an open discussion of hybrid-decision situations. In future, it might not only be important to address the technical question of what we can achieve by using computers but also how humans perceive computer actions and decisions. The research on artificial moral agency as well as how computers affect our moral considerations is just emerging.

References

Allen, C., Varner, G., and Zinser, J. (2000). Prolegomena to any future artificial moral agent. *Journal of Experimental & Theoretical Artificial Intelligence*, 12(3):251–261.

Andreoni, J. and Bernheim, B. D. (2009). Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects. *Econometrica*, 77(5):1607–1636.

Andreoni, J. and Petrie, R. (2004). Public goods experiments without confidentiality: A glimpse into fund-raising. *Journal of Public Economics*, 88(7-8):1605–1623.

Aronson, E. (2009). The Return of the Repressed: Dissonance Theory Makes a Comeback. *Psychological Inquiry*, 3(4):303–311.

Asaro, P. M. (2011). A Body to Kick, but Still No Soul to Damn: Legal Perspectives on Robotics. In Lin, K. Abney, and G. Bekey, editor, *Robot Ethics: The Ethical and Social Implications of Robotics*, pages 169–186. MIT Press, Cambridge, MA.

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

Battigalli, P. and Dufwenberg, M. (2007). Guilt in Games. *American Economic Review*, 97(2):170–176.

Beauvois, J.-L. and Joule, R. (1996). *A radical dissonance theory*. Taylor & Francis, London; Bristol, PA.

Bechel, W. (1985). Attributing Responsibility to Computer Systems. *Metaphilosophy*, 16(4):296–306.

Bénabou, R. and Tirole, J. (2006). Incentives and Prosocial Behavior. *American Economic Review*, 96(5):1652–1678.

Berndsen, M. and Manstead, A. S. R. (2007). On the relationship between responsibility and guilt: Antecedent appraisal or elaborated appraisal? *European Journal of Social Psychology*, 37(4):774–792.

Bland, J. and Nikiforakis, N. (2015). Coordination with third-party externalities. *European Economic Review*, 80:1–15.

Bodner, R. and Prelec, D. (2003). Self-signaling and diagnostic utility in everyday decision making. *The psychology of economic decisions*, 1:105–126.

Bornstein, G. and Yaniv, I. (1998). Individual and group behavior in the ultimatum game: Are groups more "rational" players? *Experimental Economics*, 1(1):101–108.

Bruun, O. and Teroni, F. (2011). Shame, Guilt and Morality. *Journal of Moral Philosophy*, 8(2):223–245.

Chen, D. L. and Schonger, M. (2013). Social Preferences or Sacred Values? Theory and Evidence of Deontological Motivations. *Working Paper, ETH Zürich, Mimeo*.

Choi, S., Thalmayr, F., Wee, D., and Weig, F. (2016). Advanced driver-assistance systems: Challenges and opportunities ahead. `http://www.mckinsey.com/industries/semiconductors/our-insights/advanced-driver-assistance-systems-challenges-and-opportunities-ahead`.

Cooper, D. J. and Kagel, J. H. (2005). Are Two Heads Better Than One? Team versus Individual Play in Signaling Games. *American Economic Review*, 95(3):477–509.

Dana, J., Weber, R. A., and Kuang, J. X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1):67–80.

Darley, J. and Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 8(4, Pt.1):377–383.

de Hooge, I. E., Nelissen, R. M. A., Breugelmans, S. M., and Zeelenberg, M. (2011). What is moral about guilt? Acting "prosocially" at the disadvantage of others. *Journal of Personality and Social Psychology*, 100(3):462–473.

DeBaets, A. M. (2014). Can a Robot Pursue the Good? Exploring Artificial Moral Agency. *Journal of Evolution and Technology*, 24:76–86.

Dennett, D. C. (1997). When HAL Kills, Whos to Blame?: Computer Ethics. *Rethinking responsibility in science and technology*, pages 203–214.

Ellingsen, T. and Johannesson, M. (2008). Pride and Prejudice: The Human Side of Incentive Theory. *American Economic Review*, 98(3):990–1008.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14(4):583–610.

Eyssel, F. and Hegel, F. (2012). (S)hes Got the Look: Gender Stereotyping of Robots. *Journal of Applied Social Psychology*, 42:2213–2230.

Falk, A. and Szech, N. (2013). Morals and Markets. *Science*, 340(6133):707–711.

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

Fischer, J. M. (1999). Recent Work on Moral Responsibility. *Ethics*, 110(1):93–139.

Fischer, P., Krueger, J. I., Greitemeyer, T., Vogrincic, C., Kastenmüller, A., Frey, D., Heene, M., Wicher, M., and Kainbacher, M. (2011). The bystander-effect: A meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin*, 137(4):517–537.

Floridi, L. and Sanders, J. W. (2004). On the Morality of Artificial Agents. *Minds and Machines*, 14(3):349–379.

Forsyth, D. R., Zyzniewski, L. E., and Giammanco, C. A. (2002). Responsibility Diffusion in Cooperative Collectives. *Personality and Social Psychology Bulletin*, 28(1):54–65.

Freeman, S., Walker, M. R., Borden, R., and Latane, B. (1975). Diffusion of Responsibility and Restaurant Tipping: Cheaper by the Bunch. *Personality and Social Psychology Bulletin*, 1(4):584–587.

Friedman (1995). "It's the Computer's Fault" –Reasoning About Computers as Moral Agents. In *Conference companion on Human factors in computing systems (CHI 95)*, pages 226–227. Association for Computing Machinery, New York, NY.

Friedman, B. and Kahn, P. H. (1992). Human agency and responsible computing: Implications for computer system design. *Journal of Systems and Software*, 17(1):7–14.

Gogoll, J. and Uhl, M. (2016). Automation and Morals — Eliciting Folk Intuitions. *TU München Peter Löscher-Stiftungslehrstuhl für Wirtschaftsethik Working Paper Series*.

Greiner, B. (2004). An online recruitment system for economic experiments. In Kremer, K. and Macho, V., editors, *Forschung und wissenschaftliches Rechnen*, pages 79–93. Göttingen.

Grossman, Z. (2014). Strategic Ignorance and the Robustness of Social Preferences. *Management Science*, 60(11):2659–2665.

Grossman, Z. (2015). Self-signaling and social-signaling in giving. *Journal of Economic Behavior & Organization*, 117:26–39.

Grossman, Z. and van der Weele, J. J. (2013). Self-Image and Willful Ignorance in Social Decisions. *Forthcoming in the Journal of the European Economic Association.*

Haidt and Kesebir (2010). Morality. In Fiske, S. T., Gilbert, D. T., Lindzey, G., and Jongsma, A. E., editors, *Handbook of social psychology*. Wiley, Hoboken, N.J.

Haisley, E. C. and Weber, R. A. (2010). Self-serving interpretations of ambiguity in other-regarding behavior. *Games and Economic Behavior*, 68(2):614–625.

Johnson, D. G. (2006). Computer systems: Moral entities but not moral agents. *Ethics and Information Technology*, 8(4):195–204.

Johnson, D. G. and Powers, T. M. (2005). Computer Systems and Responsibility: A Normative Look at Technological Complexity. *Ethics and Information Technology*, 7(2):99–107.

Katagiri, Y., Nass, C., Takeuchi, and Yugo (2001). Cross-Cultural Studies of the Computers are Social Actors Paradigma: The Case of Reciprocity. In Smith, M. J., Koubek, R. J., Salvendy, G., and Harris, D., editors, *Usability evaluation and interface design*, volume 1 of *Human factors and ergonomics*, pages 1558–1562. Lawrence Erlbaum, Mahwah, N.J. and London.

Kaushik, D., High, R., Clark, C. J., and LaGrange, C. A. (2010). Malfunction of the Da Vinci robotic system during robot-assisted laparoscopic prostatectomy: an international survey. *Journal of endourology*, 24(4):571–575.

Kocher, M. G. and Sutter, M. (2005). The Decision Maker Matters: Individual Versus Group Behaviour in Experimental Beauty-Contest Games*. *The Economic Journal*, 115(500):200–223.

Kocher, M. G. and Sutter, M. (2007). Individual versus group behavior and the role of the decision making procedure in gift-exchange experiments. *Empirica*, 34(1):63–88.

Konow, J. (2000). Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions. *American Economic Review*, 90(4):1072–1092.

Kugler, T., Bornstein, G., Kocher, M. G., and Sutter, M. (2007). Trust between individuals and groups: Groups are less trusting than individuals but just as trustworthy. *Journal of Economic Psychology*, 28(6):646–657.

Latané, B. and Nida, S. (1981). Ten years of research on group size and helping. *Psychological Bulletin*, 89(2):308–324.

Lipinski, T. A., Buchanan, E. A., and Britz, J. J. (2002). Sticks and stones and words that harm: Liability vs. responsibility, section 230 and defamatory speech in cyberspace. *Ethics and Information Technology*, 4(2):143–158.

Luhan, W., Kocher, M., and Sutter, M. (2009). Group polarization in the team dictator game reconsidered. *Experimental Economics*, 12(1):26–41.

Matthey, A. and Regner, T. (2011). Do I Really Want to Know? A Cognitive Dissonance-Based Explanation of Other-Regarding Behavior. *Games*, 2(4):114–135.

May, L. (1992). *Sharing responsibility*. University of Chicago Press, Chicago.

Mazar, N., Amir, O., and Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research*, 45(6):633–644.

McGlynn, R. P., Harding, D. J., and Cottle, J. L. (2009). Individual-Group Discontinuity in Group-Individual Interactions: Does Size Matter? *Group Processes & Intergroup Relations*, 12(1):129–143.

Melo, C. d., Marsella, S., and Gratch, J. (2016). People Do Not Feel Guilty About Exploiting Machines. *ACM Transactions on Computer-Human Interaction*, 23(2):1–17.

Moon, Y. (2003). Don't Blame the Computer: When Self-Disclosure Moderates the Self-Serving Bias. *Journal of Consumer Psychology*, 13(1-2):125–137.

Moon, Y. and Nass, C. (1998). Are computers scapegoats? Attributions of responsibility in human–computer interaction. *International Journal of Human-Computer Studies*, 49(1):79–94.

Moor, Johnson, D. G., and Snapper, J. W. (1985). Are there decisions computers should never make? In Maner, W., Johnson, D. G., and Snapper, J. W., editors, *Ethical issues in the use of computers*, pages 120–130. Wadsworth Publ. Co., Belmont, CA.

Moore, M. S. (1999). Causation and Responsibility. *Social Philosophy and Policy*, 16(2):1–51.

Murnighan, J., Oesch, J. M., and Pillutla, M. (2001). Player Types and Self-Impression Management in Dictatorship Games: Two Experiments. *Games and Economic Behavior*, 37(2):388–414.

Nass, C., Fogg, B. J., and Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, 45(6):669–678.

Nass, C. and Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56(1):81–103.

Nass, C., Steuer, J., and Tauber, E. R. (1994). Computers are social actors. In Adelson, B., Dumais, S., and Olson, J., editors, *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 72–78, New York. Association for Computing Machinery.

Nissenbaum, H. (1994). Computing and accountability. *Communications of the ACM*, 37(1):72–80.

Panchanathan, K., Frankenhuis, W. E., and Silk, J. B. (2013). The bystander effect in an N-person dictator game. *Organizational Behavior and Human Decision Processes*, 120(2):285–297.

R Development Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rabin, M. (1995). Moral Preferences, Moral Constraints, and Self-Serving Biases. *Department of Economics UCB (unpublished manuscript)*.

Reeves, B. and Nass, C. (2003). *The media equation: How people treat computers, television, and new media like real people and places*. CSLI Publ, Stanford CA, 1. paperback ed., 3. [print.] edition.

Rockenbach, B., Sadrieh, A., and Mathauschek, B. (2007). Teams take the better risks. *Journal of Economic Behavior & Organization*, 63(3):412–422.

Rothenhäusler, D., Schweizer, N., and Szech, N. (2013). Institutions, Shared Guilt, and Moral Transgression. *Working Paper Series in Economics*, (47).

Scheines, R. (2002). Computation and Causation. *Metaphilosophy*, 33(1/2):158–180.

Seaman, A. M. (2016). Completely automated robotic surgery: on the horizon? `http://www.reuters.com/article/us-health-surgery-robot-idUSKCN0Y12Q2`.

Senthilingam, M. (2016). Would you let a robot perform your surgery by itself? `http://edition.cnn.com/2016/05/12/health/robot-surgeon-bowel-operation/`.

Sherman, N. (1999). Taking Responsibility for our Emotions. *Social Philosophy and Policy*, 16(02):294–323.

Snapper, J. W. (1985). Responsibility for computer–based errors. *Metaphilosophy*, 16(4):289–295.

Sparrow, R. (2007). Killer Robots. *Journal of Applied Philosophy*, 24(1):62–77.

Stice, E. (1992). The similarities between cognitive dissonance and guilt: Confession as a relief of dissonance. *Current Psychology*, 11(1):69–77.

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, Greg, Hirschberg, Julia, Kalyanakrishnan, S., Kamar, E., Kraus, S., Leyton-Brown, K., Parkes, D., Press, W., Saxenian, A., Shah, J., Tambe, M., and Teller, A. (2016). Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial In-telligence: Report of the 2015-2016 Study Panel. `http://ai100.stanford.edu/2016-report`.

Sullins, J. P. (2006). When Is a Robot a Moral Agent? In Adelson, M. and Anderson, S., editors, *Machine Ethics*, pages 151–160, New York, NY. Association for Computing Machinery.

Sutter, M. (2005). Are four heads better than two? An experimental beauty-contest game with teams of different size. *Economics Letters*, 88(1):41–46.

Wallace, R. J. (1994). *Responsibility and the moral sentiments*. Harvard University Press, Cambridge, Mass.

Wallach, M. A., Kogan, N., and Bem, D. J. (1964). Diffusion of responsibility and level of risk taking in groups. *The Journal of Abnormal and Social Psychology*, 68(3):263–274.

Wildschut, T., Pinter, B., Vevea, J. L., Insko, C. A., and Schopler, J. (2003). Beyond the group mind: a quantitative review of the interindividual-intergroup discontinuity effect. *Psychological Bulletin*, 129(5):698–722.

A. Appendix for Online Publication

In addition to the data used to test your hypotheses we collected some further data which we provide here. We also present some additional information on the interfaces and questions used. Data and Methods can be found at `https://www.kirchkamp.de/research/shareMachine.html`.

A.1. Interfaces and Questions

In this section the interfaces as well as the questions used in the experiment are presented.

A.1.1. Dictator Game Interface

In the MDT as well as in the CDT dictators used the interface sketched in Figure 11 to enter their decision. Responders used the interface sketched in Figure 12 to enter their guess.

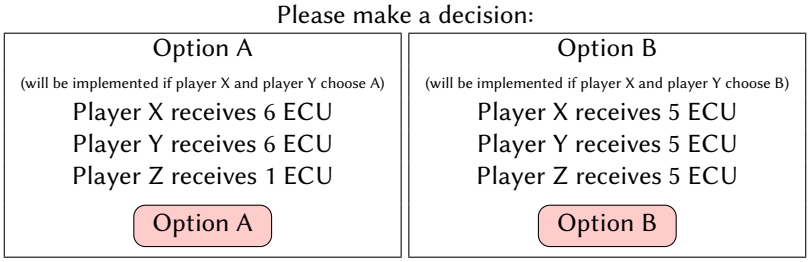

Figure 11: Dictator Game interface for dictators

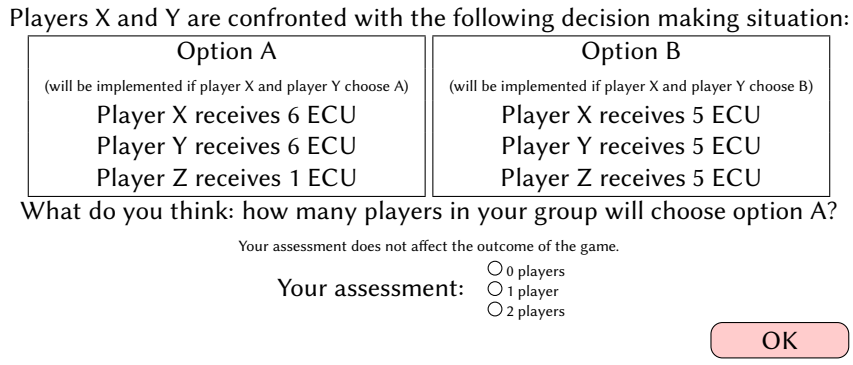

Figure 12: Dictator Game interface for responders

The interfaces for dictators and responders in the SDT were similar to the interfaces used in the MDT and in the CDT with the exception that there were just two players, of which one was a dictator who would gain an advantage if Option A becomes implemented. Accordingly, responder in the SDT were just asked regarding their guess for Player X's choice.

A.1.2. Questions

To determine the levels of responsibility and guilt that the dictators perceived all subjects were asked to answer some questions. The questions were asked right after the decision and before the final outcome and payoff was announced. They differed slightly between the treatments as they were adjusted to the different situations. The questions used in the MDT for the subject in the roll of Player X are presented below as an example. The used answer method is presented in brackets. The same questions were asked in the CDT, however *"Player Y"* was replaced by *"the computer"*. The questions were also asked in the SDT, except for the first three question. The questions for Player Y in the MDT were very similar to the questions asked to Player X. In the SDT the questions were altered as Player Y did not decide on her own. In the SDT the questions for Player Y were similar to the questions asked to Player Z in the CDT and MDT as all of them were responders. Responders were asked what they expect Player X to do.

While dictators were asked directly responders and passive dictators were asked indirectly. For example, responders and passive dictators, were asked how responsible and guilty they perceive the dictator(s) and what they expect the dictator(s) would do in a specific situation. We also asked the responders and passive dictators to estimate how responsible and guilty the dictators perceive themselves for a decision in the experiment as well as in the manipulation check.

1. How would you have decided, if you would have made the decision on your own? [Slider from *"Option A"* to *"Option B"*] (for an analysis of the answers given see Sections A.2.1, A.4.1, A.6.1)

2. What likelihood did you assume for Player Y to choose Option A (Player X receives 6 ECU, Player Y receives 6 ECU, Player Z receives 1 ECU)? [Slider from *"Player Y always chooses A"* to *"Player Y always chooses B"*] (for an analysis of the answers given see Sections A.2.2, A.4.2, A.6.2)

3. Did the likelihood you assumed for Player Y to choose Option A (Player X receives 6 ECU, Player Y receives 6 ECU, Player Z receives 1 ECU) affect your decision? [*Radio buttons "YES"; "NO"*] (for an analysis of the answers given see Sections A.2.3)

4. Why did you choose Option A (Player X receives 6 ECU, Player Y receives 6 ECU, Player Z receives 1 ECU)? [Open question with a maximum of 100 characters] / Why did you choose Option B (Player X receives 5 ECU, Player Y receives 5 ECU, Player Z receives 5 ECU)? [Open question with a maximum of 100 characters] (for the answers given see online dataset)

5. What could be additional reasons for choosing option A(player X receives 6 ECU, player Y receives 6 ECU, player Z receives 1 ECU)? [Open question with a maximum of 100 characters] (for the answers given see online dataset)

6. I feel responsible for the payoff of Player Z. [Slider from *"Very responsible"* to *"Not responsible at all"*] (for an analysis of the answers given see Sections 5.1, A.4.3, A.6.3)

7. I feel responsible for the payoff of Player Y. [Slider from *"Very responsible"* to *"Not responsible at all"*] (for an analysis of the answers given see Sections 5.1, A.4.3, A.6.3)

8. Option A will be implemented if you and the other player chose A. In this case, Player X receives 6 ECU, Player Y receives 6 ECU and Player Z receives 1 ECU. Do you feel gulity in this case? [Slider from *"I feel very guilty"* to *"I do not feel guilty at all"*] (for an analysis of the answers given see Sections 5.2, A.6.5)

9. Option A will be implemented if you and the other player chose A. In this case, Player X receives 6 ECU, Player Y receives 6 ECU and Player Z receives 1 ECU. Please adjust the slide control, so that it shows how you perceive your responsibility as well as the responsiility of the other player if option A is implemented. [Slider from *"I am fully responsible"* to *"I am not responsible"* and slider from *"My fellow player is fully responsible"* to *"My fellow player is not responsible"*] (for an analysis of the answers given see Sections 5.1, A.4.3, A.6.3)

In addition to these questions, a manipulation check was conducted in all treatments. Subjects participating in the MDT were asked in the manipulation check how responsible and guilty they would feel for the final payoff if the decision of one of the dictators would be made by a computer instead of a human.[30] Subjects participating in the CDT were asked in the manipulation check how responsible and guilty they would feel for the final decision in a situation similar to the MDT (two humans decide together which option should be implemented instead of a computer).[31] In the SDT subjects were asked in the manipulation check how responsible and guilty they would feel for the final decision if a computer instead of themselves would decide which option will be implemented.[32] As an example, the questions for Player X used in the MDT manipulation check are presented below.

1. How responsible would you feel in this situation for the payoff of Player Y? [Radio buttons *"As responsible as in the experiment"*; *"More responsible than in the experiment"*; *"Less responsible than in the experiment"*] (for an analysis of the answers given see Sections A.3.2, A.5.2, A.7.2)

[30] The wording of the manipulation check in the MDT was *"Imagine, now the decision of player X [Y] is made by a computer. The likelihood the computer chooses Option A (Player X receives 6 ECU, Player Y receives 6 ECU and Player Z receives 1 ECU) or Option B (Player X receives 5 ECU, Player Y receives 5 ECU and Player Z receives 1 ECU) is as high as the likelihood experimental subjects chose Option A or Option B in a former experiment. Example: If three out of ten participants in a former experiment, whose decision affected the payment, chose a particular option the computer would choose that option with a probability of 30%. The participants in the former experiment were not told that their decision would affect a computer's decision in this experiment. Please compare this decision-making situation with the one Player X and Player Y are confronted with in this experiment."*.

[31] The corresponding wording of the manipulation check in the CDT was *"Imagine, now the decision would not be made by a computer but by player Y[X] him/herself. Please compare this decision situation to the situation you were confronted with in this experiment."*.

[32] The adjusted first sentence of the manipulation check in SDT was *"Imagine, now the decision of player X is made by a computer."*

2. How responsible would you feel in this situation for the payoff of Player Z? [Radio buttons *"As responsible as in the experiment"*; *"More responsible than in the experiment"*; *"Less responsible than in the experiment"*] (for an analysis of the answers given see Sections A.3.3, A.5.3, A.7.3)

3. How guilty would you feel if you and the computer both chose Option A and therefore Option A (Player X receives 6 ECU, Player Y receives 6 ECU, Player Z receives 1 ECU) would be implemented? [Radio buttons *"As guilty as in the experiment"*; *"More guilty than in the experiment"*; *"Less guilty than in the experiment"*] (for an analysis of the answers given see Sections A.3.4, A.5.4, A.7.4)

4. Option A will be implemented if you and the computer chose Option A. In this case, Player X receives 6 ECU, Player Y receives 6 ECU and Player Z receives 1 ECU. Please adjust the slide control, so that it shows your preceived responsibility as well as the responsibility you allocate to the computer if option A is implemented.[Slider from *"I am responsible"* to *"I am not responsible"* and slider from *"The computer is fully responsible"* to *"The computer is not responsible"*] (for an analysis of the answers given see Sections A.3.1, A.5.1, A.7.1)

A.2. Dictator: Further Measurements

In addition to question regarding the perceived responsibility and guilt for the outcome we asked the dictators further questions about their expectations. Even if these questions are not necessary to our research question the results may be interesting for others.

A.2.1. Deciding Alone

Subjects were able to insert their assessment by a continuous scale from *"Option A"* (0) to *"Option B"* (100). A large proportion of the actively deciding dictators in the CDT and in the MDT who had chosen Option A stated that they would have chosen Option A if they would have had to decide on their own as Figure 13 shows. This was stronger for dictators in the CDT than for dictators in the MDT (p-value 0.0000). In accordance, dictators in the MDT as well as in the CDT who had chosen Option B stated a higher probability of choosing Option B if they would have had to decide alone.

A.2.2. Expectation Regarding the Behavior of the Other Human Dictator or Computer

The expectation was measured by a continuous scale from *"Player [Computer] choose always A"* (0) to *"Player [Computer] choose always B"* (100). Dictators expected the other eighter human or computer dictator to make a choice similar to their own as Figure 14. Interesingly, dictators in the MDT expected the other human dictator to choose Option A significantly more often than dictators in the CDT expected the computer to choose Option A (p-value 0.0011). This result was mainly driven by dictators in the MDT who had chosen Option A (p-value 0.0001).

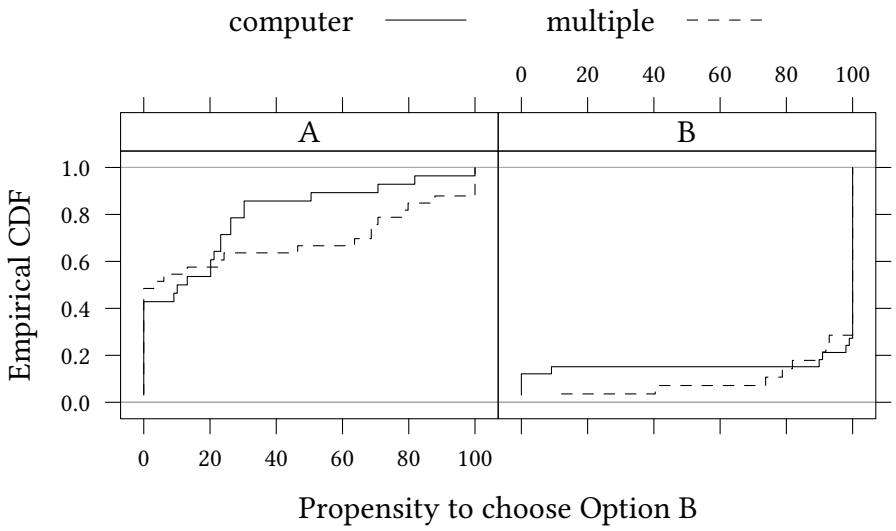

Figure 13: Deciding alone (as a hypothetical single dictator) as seen by actively deciding dictators
(Question 1 from Section A.1.2)

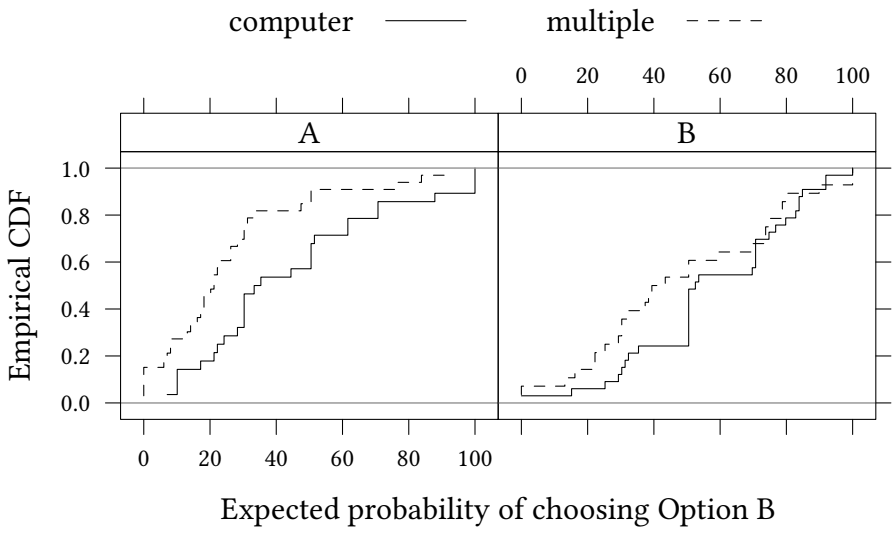

Figure 14: Expected other dictators' choice as seen by actively deciding dictators
(Question 2 from Section A.1.2)

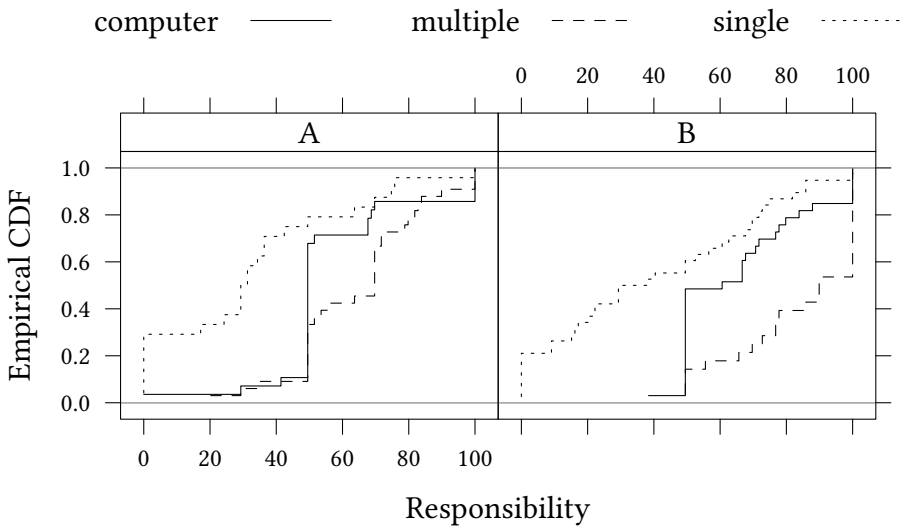

Figure 15: Perceived own responsibility in the manipulation check as seen by dictators (Question 4 from Section A.1.2)

A.2.3. Influence on Own Decision

Dictators could either choose *"YES"* or *"NO"*. In both treatments, the MDT and the CDT, the fraction of dictators who stated that they took the expected decision of the other decider into account when making their own decision was very similar (34.4% in the MDT and by 36.1% in the CDT).

A.3. Dictator: Manipulation Check

Dictators participating in the MDT [CDT] were asked to state the responsibility and guilt they would perceive for the outcome as well as the level of responsibility they would allocate to the other dictator if, contrary to the game they just played, they would have to decide together with a computer [another human dictator]. Dictators participating in the SDT were asked to state how responsible they would feel for the outcome if a computer would decide instead of themselves. [33]

A.3.1. Manipulation Check: Perceived Own and Others Responsibility

The perceived responsibility was measured by a continuous scale from *"Not responsible at all"* (0) to *"Very responsible"* (100). The own responsibility perceived by the dictators in the manipulation check is shown in Figure 15. Dictators in the SDT stated to perceive themselves to be not very responsible if the decision would be made by a computer. Interestingly, dictators in the CDT stated to perceive themselves to be less responsible for the final payoff

[33] A detailed description of the manipultation check can be found in Section A.1.2.

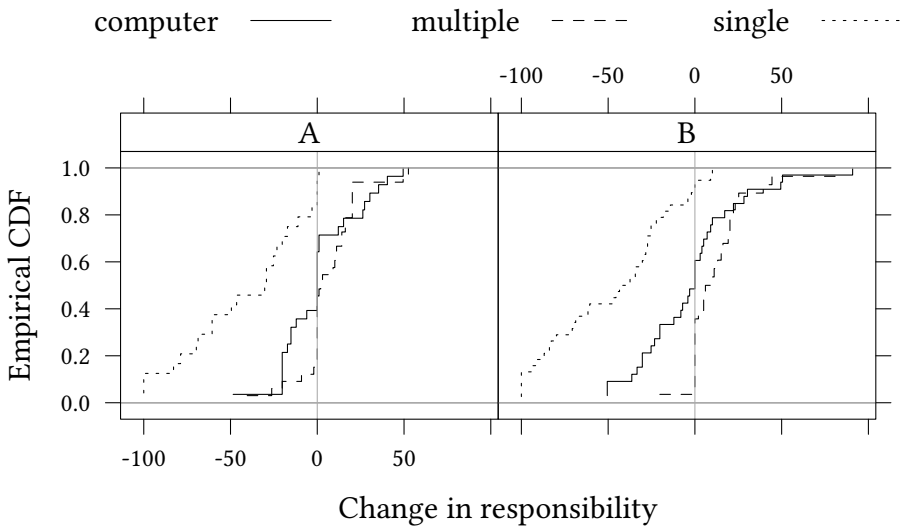

Figure 16: Change in own responsibility in the manipulation check as seen by dictators
The figure shows the difference between the own responsibility in the hypothetical situation described in Section A.3 and the own responsibility in the actual experiment as perceived by the dictators (as shown in Figure 1).

if they would have to decide with another human dictator, than dictators in the MDT, who would have to decide with a computer instead of another human dictator (p-value 0.0011).

For a comparison of the relative change between the perceived own responsibility in the hypothetical situation and the perceived own responsibility in the actual experiment see Figure 16. In line with Hypothesis 1.i , dictators in the SDT stated that they would feel less responsible if a computer would decide on their behalf (p-value 0.0000). In line with Hypothesis 1.ii, the perceived own responsibility also increased for dictators in the MDT when their counterpart would be replaced by a computer (p-value 0.0130). Interestingly, the perceived own responsibility did not decrease significantly for dictators in the CDT when their counterpart would be replaced by a human (p-value 0.5806).

The responsibility allocated to the other dictator by the dictators in the manipulation check is shown in Figure 17. Significantly more responsibility was allocated to a potential human dictator in the CDT maipulation check compared to a potential computer dictator in the MDT (p-value 0.0001).

The increase or decrease in the responsibility allocated to the other dictator between the hypothetical situation and the actual experiment is shown in Figure 18. In line with Hypothesis 1.ii, responsibility attributed to the other player in the CDT increases significantly once the other player is no longer a computer (p-value 0.0196). Similarly, responsibility decreases significantly in the MDT once the other player is no longer a human (p-value 0.0000).

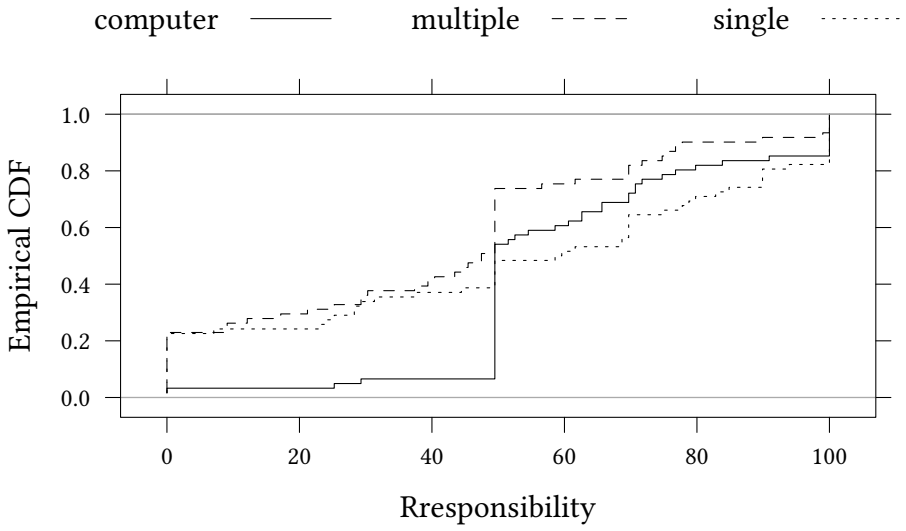

Figure 17: Allocated responsibility to the other eighter human or computer dictator in the manipulation check as seen by dictators
(Question 4 from Section A.1.2)

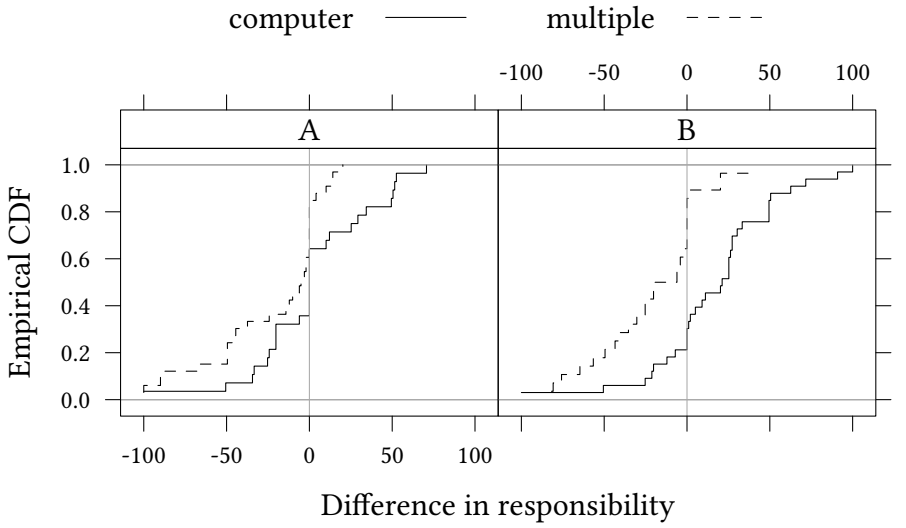

Figure 18: Difference in responsibility allocated to the human or computer dictator in the manipulation check as seen by dictators
The Figure shows the difference in the responsibility allocated by the dictator to the other eighter human or computer dictator between the hypothetical situation (described in Section A.3) and the actual experiment (as shown in Figure 3).

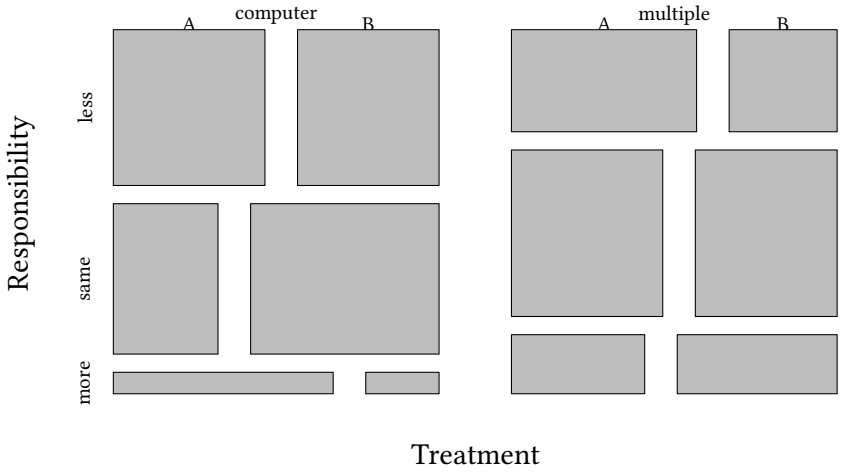

Figure 19: Change in responsibility for the other eighter active or passive dictator in the manipulation check as seen by dictators
(Question 1 from Section A.1.2)

A.3.2. Manipulation Check: Responsibility for the Other Dictator

The perceived responsibility was measured by using three statements: *"Same level of responsibility as in the experiment before"*, *"More responsible as in the experiment before"* and *"Less responsible as in the experiment before"*. Results are shown in Figure 19. Hypothesis 1.ii suggests that dictators in the CDT who would share their decision with a human instead of a computer would feel less responsible for the payoff of the other dictator than before. This is confirmed by a binomial test (p-value from a binomial test 0.0000). Similarly we expect that dictators in the MDT who would share their decision with a computer feel more responsible for the payoff of the other dictator. However, this was not the case (p-value from a binomial test 0.2005).

A.3.3. Manipulation Check: Responsibility for the Responder

The perceived responsibility was measured by using three statements: *"Same level of responsibility as in the experiment before"*, *"More responsible as in the experiment before"* and *"Less responsible as in the experiment before"*. Details are shown in Figure 20. In line with Hypothesis 1.i we expected the dictators in the SDT to feel less responsibility if the decision would be made by a computer and not by themselves. As we see in Figure 20, these dictators indeed felt significantly less responsibility for the payoff of the responder (p-value from a binomial test 0.0000). In line with Hypothesis 1.ii we expected dictators in the CDT to feel less responsible once they can share the burden of their choice with a human instead of a computer and vice versa. Again, this was confirmed by the results (see Figure 20) (p-value from a binomial test 0.0009). Similarly, we expected dictators in the MDT to feel more responsible for the payoff of the responder once their human counterpart is replaced with a computer. The effect,

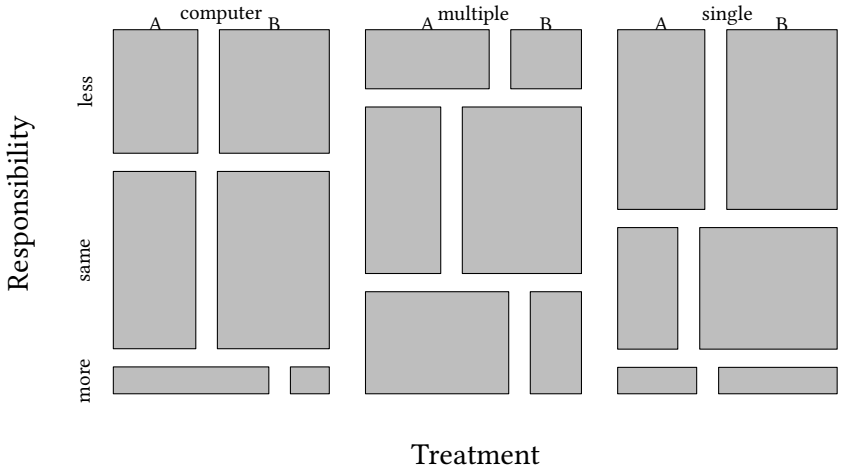

Figure 20: Change in responsibility for the responder in the manipulation check as seen by dictators
(Question 2 from Section A.1.2)

however, was only small and not significant (p-value from a binomial test 0.2005).

A.3.4. Manipulation Check: Perceived Guilt

The perceived guilt was measured as *"Same level of guilt as in the experiment before"*, *"More guilt as in the experiment before"* and *"Less guilt as in the experiment before"*. Details are shown in Figure 21. In line with Hypothesis 2.i we expected dictators in the SDT to feel less guilty when the actual decision is taken by a computer and not by the dictators. As we see in Figure 21 the dictators felt indeed significantly less guilty (p-value from a binomial test 0.0000).

In line with Hypothesis 2.ii we did expect dictators in the CDT to feel less guilty once they can share the burden of their choice with a human instead of a computer and vice versa. Figure 21 shows such a tendency, but the effect is not significant (p-value from a binomial test 0.3269). Similarly, we expected dictators in the MDT to feel more guilty once their human counterpart is replaced with a computer. However, this effect was also not significant (p-value from a binomial test 0.0963).

A.4. Responder: Further Measurements

Similar to the questions for the dictator(s) presented in Section A.1.2 we asked the responders in all treatments about their expectations regarding the dictators' behavior and perception of responsibility and guilt. Even if these questions are not needed to answer our research questions the results might be interesting for others.

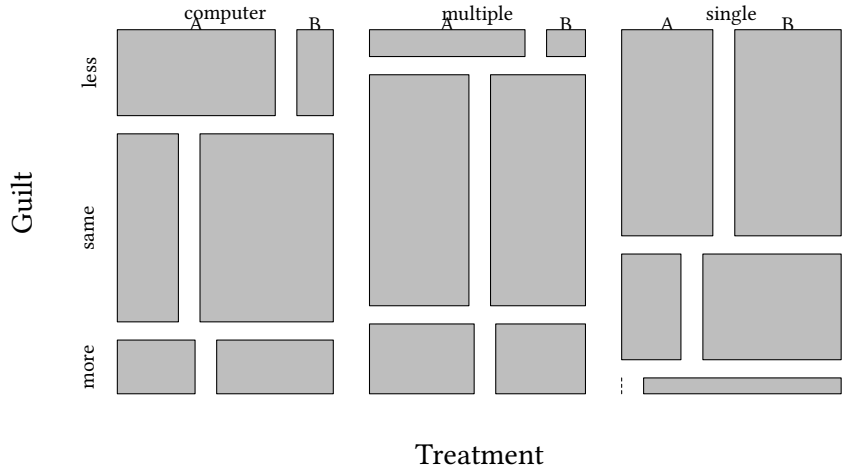

Figure 21: Change in guilt in the manipulation check as seen by dictators
(Question 3 from Section A.1.2)

A.4.1. Deciding Alone

Responders were able to insert their assessment by a continuous scale from *"Option A"* (0) to *"Option B"* (100). A large proportion of the responders in the CDT stated that they would expect that the dictators choose Option A if they he would have had to decide on their own (see Figure 22). This result was even slightly stronger in the MDT.

A.4.2. Expectation Regarding the Behavior of the Human Dictator(s) or Computer

The expectation was measured by a continuous scale from *"Player choose always Option A"* (0) to *"Player choose always Option B"* (100). The result is shown in Figure 23. Responders in the SDT expected that the dictators choose Option B with a higher probability than responders in the MDT (p-value 0.0006). Furthermore, the responders' expectation regarding the choice of the human dictators in the MDT and in the CDT did not differ significantly (p-value 0.2191).

A.4.3. Allocated Responsibility for the Other Dictator and Responder

The assigned responsibility was measured by a continuous scale from *"not responsible at all"* (0) to *"totally responsible"* (100). Figure 24 shows, responders did not expect that the dictator would perceive to be more or less responsible for the payoff of the other benefiting dictator in the MDT, where the other benefiting dictator decided on her on, than in the CDT, where the decision of the other benefiting dictator was made by a computer (p-value 0.3505).

However, as Figure 25 shows, responders perceptions regarding the responsibility of the dictator for the responders' payoff differed between the treatments. While the allocated responsibility did not differ significantly between the SDT and the MDT (p-value 0.2614), responders perceived the dictator to be significantly less responsible for their payoff in the

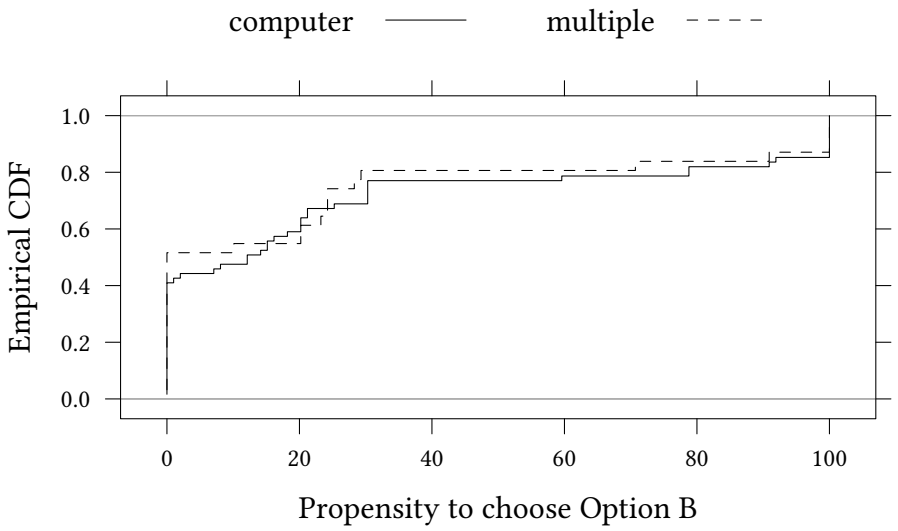

Figure 22: Expectation of responders for dictators who are deciding alone (as a hypothetical single player)
(Question 1 from Section A.1.2)

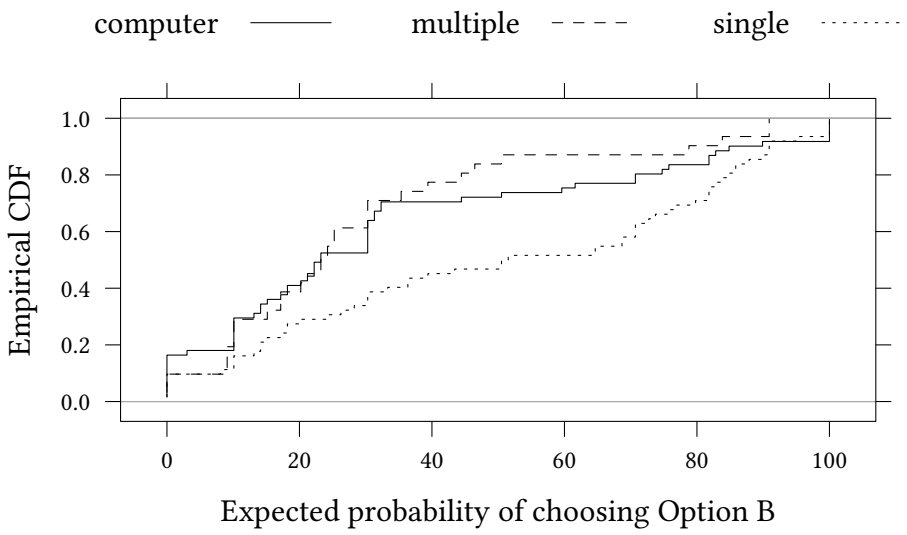

Figure 23: Expected human dictators' choice as seen by responders
(Question 2 from Section A.1.2)

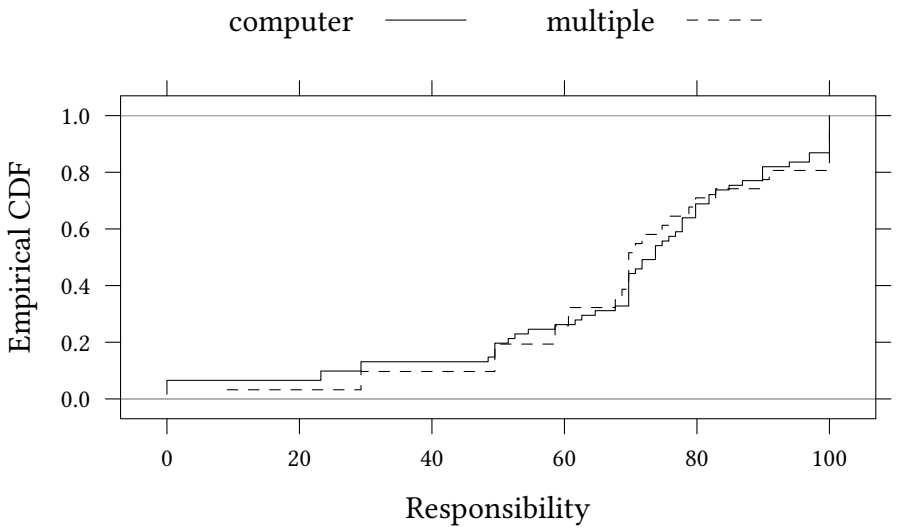

Figure 24: Allocated responsibility to the dictator(s) for the other either active or passive dictator as seen by responders
(Question 7 from Section A.1.2)

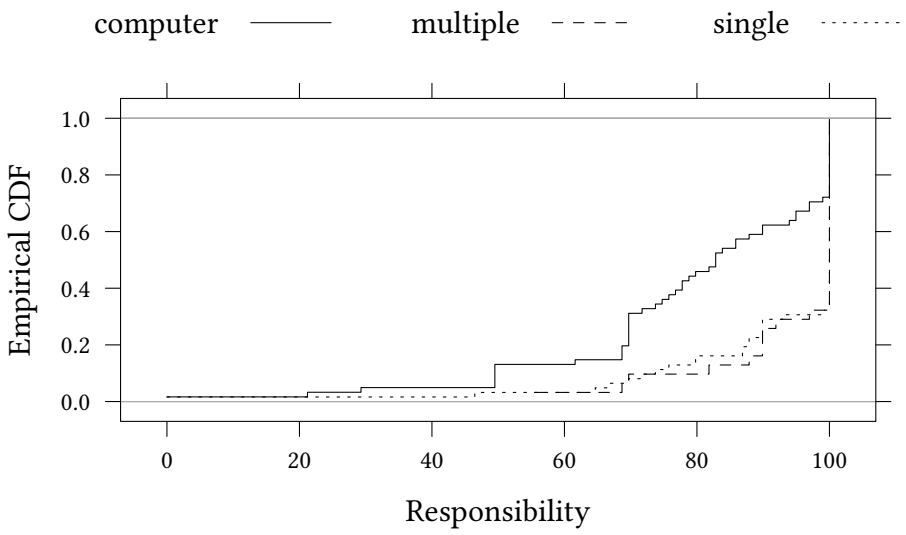

Figure 25: Allocated responsibility to the dictator(s) for the responder as seen by responders
(Question 6 from Section A.1.2)

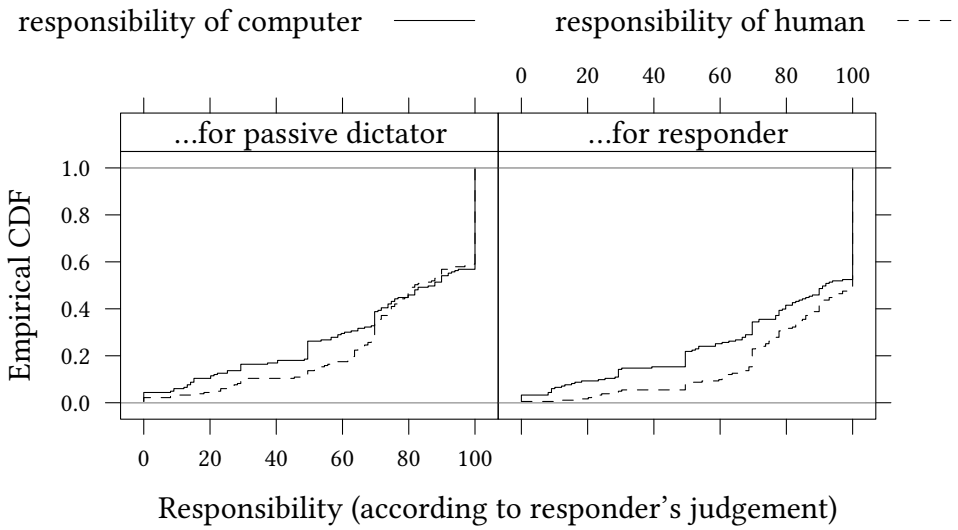

Figure 26: Allocated responsibility to the dictator(s) for the responder and the passive dictator as seen by responders

CDT than in the MDT (p-value 0.0000).

In addition, responder in the CDT stated that they perceive a human dictator to be more responsible for the final payoff of the passive dictator as well as for the payoff of the responder than a computer dictator, as shown in Figure 26.

A.5. Responder: Manipulation Check

Responders participating in the MDT [CDT] were asked to state how responsible and guilty dictators might perceive themselves to be for the final outcome if, contrary to the game they just played, they would have to decide together with a computer [another human]. Responders participating in SDT were asked to state how responsible they expect the dictators to perceive themselves for the final outcome, if a computer would decide on their behalf. [34]

A.5.1. Manipulation Check: Allocated Responsibility to Human Dictator(s) and Computer

The responsibility of the dictator for the final payoff as perceived by the responders in the manipulation check is shown in Figure 27. Responders in the SDT perceived the dictator to be less responsible if the decision would be made by a computer. Interestingly, responders also perceived the dictator in the CDT to be significantly less responsible for the final payoff if she would have to decide together with another human compared to dictators in the MDT, who would have to decide with a computer instead of another human (p-value 0.0149). For a comparison of the relative changes in the responders' perception of the responsibility of

[34] For the wording of the manipulation check see Section A.1.2. It was the same as for the dictators.

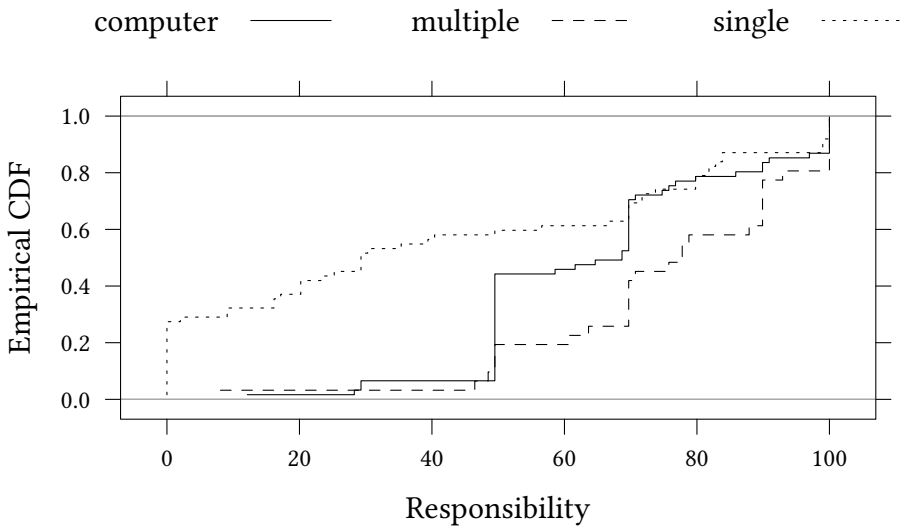

Figure 27: Expected human dictators' responsibility in the manipulation check as seen by responders
(Question 4 from Section A.1.2)

the dictator(s) for their own payoff in the hypothetical situation in the actual experiment by the responder see Figure 28. In line with Hypothesis 1.i , responders in the SDT expected the dictator to feel less responsible if a computer would decide on her behalf (p-value 0.0000). However, contrary to Hypothesis 1.ii, responders did not expect the dictators to feel significantly more responsible in the MDT when their counterpart would be replaced by a computer (p-value 0.5205). The same applies for the CDT where responders did not expect the dictators to feel responsible if their counterpart would be replaced by a human (p-value 0.1527).

The responsibility allocated in the manipulation check to the other dictator (either human or computer) perceived by the responders is shown in Figure 29. Responders in the SDT perceived the computer as significantly more responsible than the responder in the MDT (p-value 0.0031). As expected, responders perceived the human dictator in the CDT to be more responsible for the final payoff than the computer dictator in the MDT (p-value 0.0001).

For a comparison of the relative change in the responsibility of the other dictator(s) between the hypothetical situation and the actual experiment as perceived by the responders see Figure 30. Responders in the MDT would perceive a computer dictator to be less responsible than a human dictator (p-value 0.0000). Correspondingly, responder in the CDT also would perceive a human dictator to be significantly more responsible than a computer dictator (p-value 0.0242).

A.5.2. Manipulation Check: Responsibility for the Other Dictator

The perceived responsibility was measured from *"Same level of responsibility as in the experiment before"*, *"More responsible as in the experiment before"* and *"Less responsible as in the*

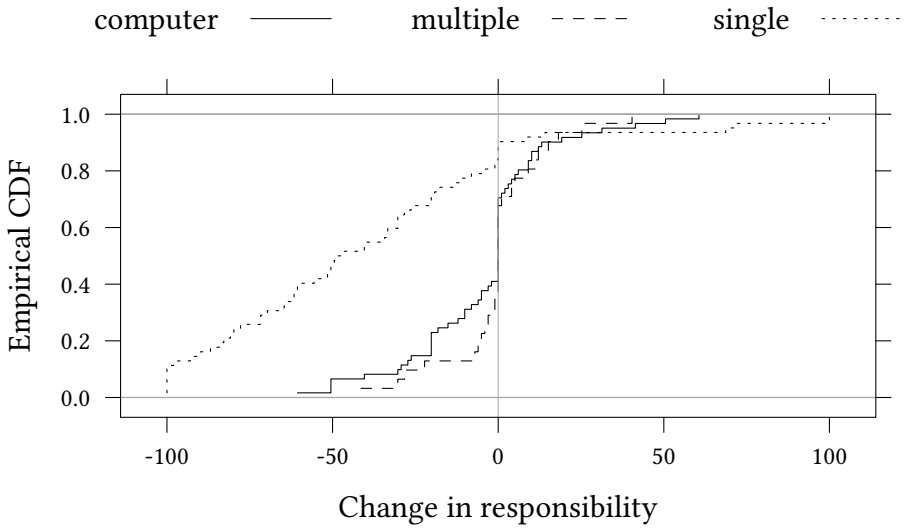

Figure 28: Change in allocated responsibility to the human dictator in the manipulation check as seen by responders

The Figure shows the difference in the responsibility that the responders expect the dictator(s) to perceive for their decision between the hypothetical situation (described in Section A.3) and the own responsibility in the actual experiment (as shown in Figure 6).

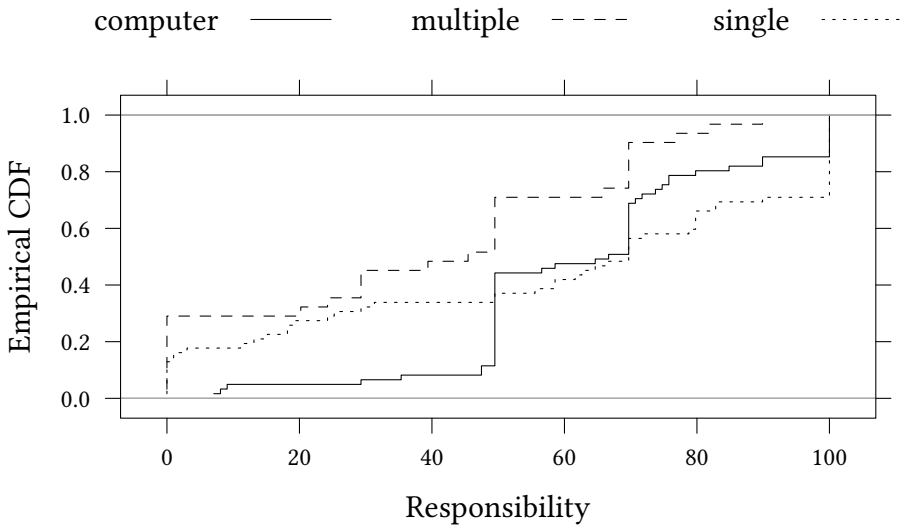

Figure 29: Expected responsibility allocated to the other eighter human or computer dictator in the manipulation check as seen by responders

(Question 4 from Section A.1.2)

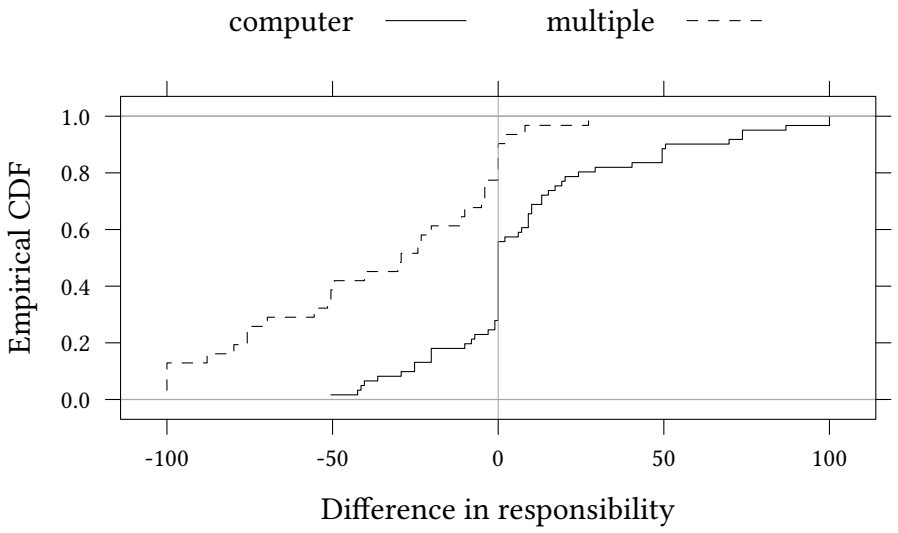

Figure 30: Difference in responsibility allocated to the human or computer dictator in the manipulation check as seen by responders

The Figure shows the difference in the responsibility allocated by the responders to the other eigther human or computer dictator between the hypothetical situation (described in Section A.3) and the actual experiment (as shown in Figure 6).

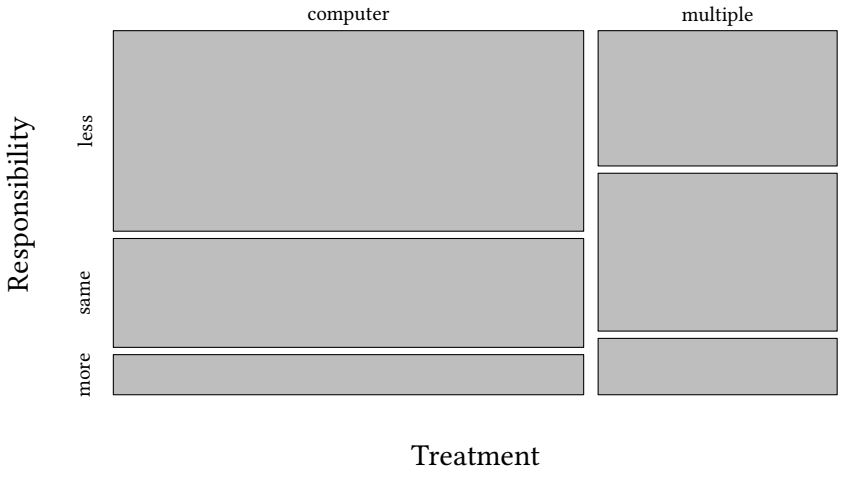

Figure 31: Change in responsibility perceived by the dictator(s) for the other eighter active or passive dictator in the manipulation check as seen by responders

(Question 2 from Section A.1.2)

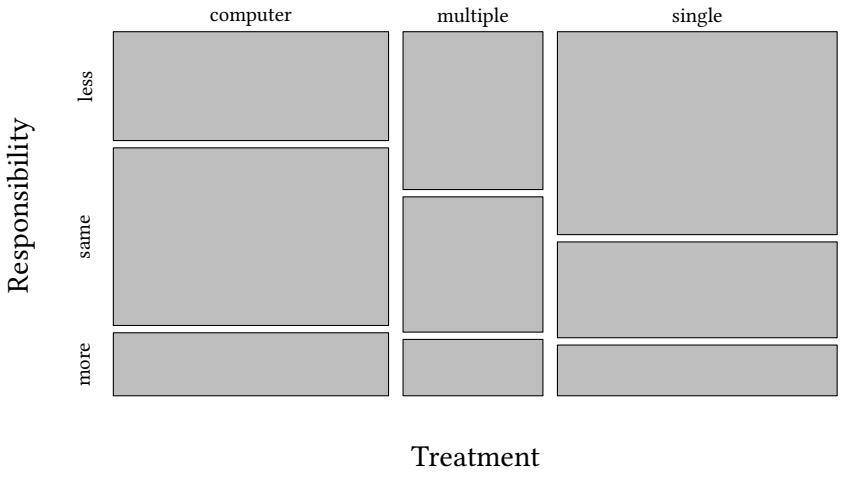

Figure 32: Change in responsibility perceived by the dictator(s) for the responder in the manipulation check as seen by responders
(Question 2 from Section A.1.2)

experiment before". Details are shown in Figure 31. Hypothesis 1.ii suggests that dictators in the CDT who are confronted with a situation where they have to share their decision with a human instead of a computer would feel less responsible than before. This is also expected by responders as Figure 31 shows (p-value from a binomial test 0.0000). Similarly we expected dictators in the MDT who would have to share their decision with a computer would feel more responsible. However, this can not be confirmed based on the answers of the responders (p-value 0.1435).

A.5.3. Manipulation Check: Responsibility for the Responder

The perceived responsibility was measured from *"Same level of responsibility as in the experiment before"*, *"More responsible as in the experiment before"* and *"Less responsible as in the experiment before"*. In line with Hypothesis 1.i we expected dictators in the SDT to feel less responsibility when the decision is taken by a computer and not by the player herself. Responders also expected that the dictator would feel significantly less responsibility as Figure 32 (p-value from a binomial test 0.0001).

Hypothesis 1.ii suggests that dictators in the CDT who would have to share their decision with a human instead of a computer would feel less responsible than before. However, this was not expected by responders (p-value from a binomial test 0.2005). Similarly we expected that dictators in the MDT who now share their decision with a computer instead of another human would feel more responsible. This was also not expected by the responders (p-value 0.0636).

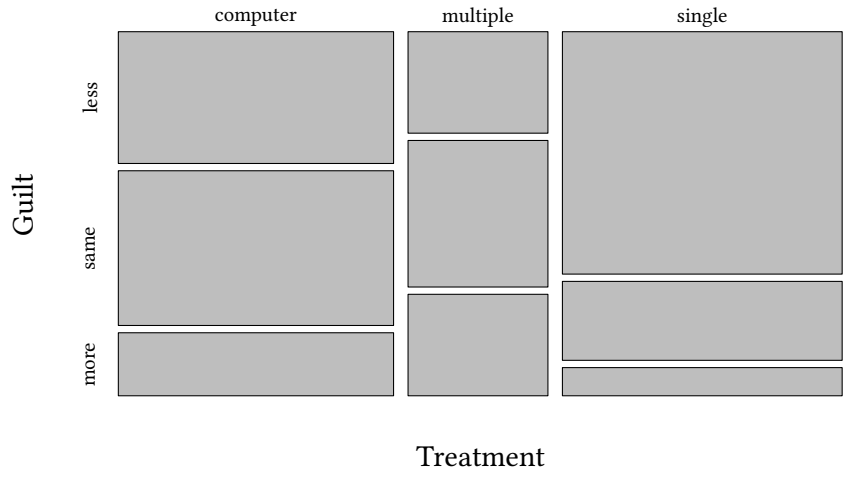

Figure 33: Change in guilt perceived by the dictator(s) in the manipulation check as seen by responder
(Question 3 from Section A.1.2)

A.5.4. Manipulation Check: Perceived Guilt

The perceived guilt was measured as *"Same level of guilt as in the experiment before"*, *"More guilt as in the experiment before"* and *"Less guilt as in the experiment before"*. In line with Hypothesis 2.i we expected dictators in the SDT to feel less guilty when the decision is taken by a computer and not by the dictator herself. This was also expected by the responders as Figure 33 shows (p-value from a binomial test 0.0000).

In line with Hypothesis 2.ii we expected dictators in the CDT to feel less guilty once they can share the burden of their choice with a human. Figure 33 shows that such a tendency was also expected by the responders, but the effect is not significant (p-value from a binomial test 0.0576). Similarly, we expected dictators in the MDT to feel more guilty once their human counterpart is replaced with a computer. However, this was not expected by the responders (p-value from a binomial test 1.0000).

A.6. Passive dictator: Further Measurements

Similar to the questions for the dictators presented in Section A.1.2 we asked the passive dictators in the CDT about their expectations regarding the dictators' behavior and perception of responsibility and guilt. Even if these questions are not necessary to our research question the results may be interesting for others.

A.6.1. Deciding Alone

Passive dictators were able to insert their assessment by a continuous scale from *"Option A"* (0) to *"Option B"* (100). A large proportion of the passive dictators stated that they expect

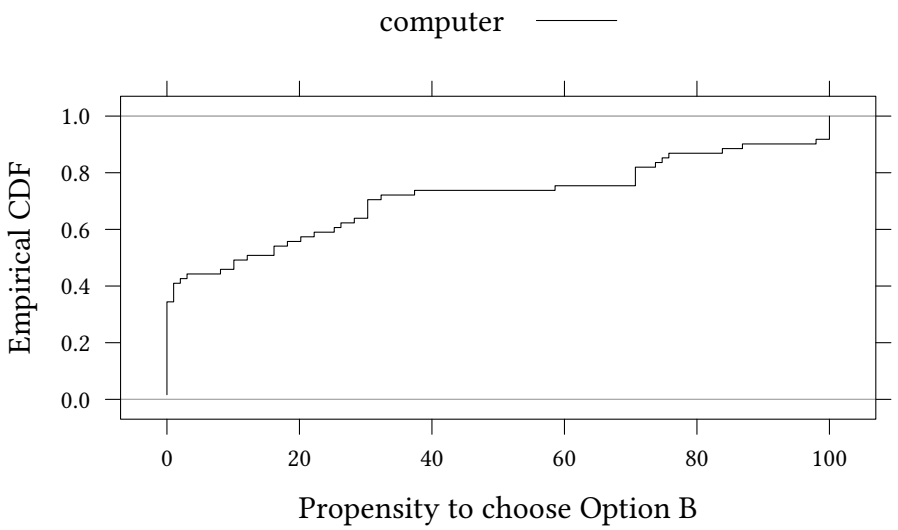

Figure 34: Expectation of passive dictators for dictators who are deciding alone (as a hypothetical single player)
(Question 1 from Section A.1.2)

that the dictators choose Option A if they would have to decide on their own as Figure 34 shows.

A.6.2. Expectation Regarding the Behavior of the Dictator(s)

The expectation was measured by a continuous scale from *"Player choose always A"* (0) to *"Player choose always B"* (100). A large proportion of the passive dictators stated that they would expect the dictator to choose Option A as Figure 35 shows.

A.6.3. Allocated Responsibility for the Decision to the Dictator(s) and the Computer

The assigned responsibility was measured by a continuous scale from *"not responsible at all"* (0) to *"totally responsible"* (100). Figure 36 shows how responsible the passive dictators perceived the dictator to be for the final outcome in the CDT. A large proportion of the passive dictators perceived the dictator to be very responsible for the final decision.

Figure 37 shows how responsible the passive dictators perceived the computer to be for the final decision in the CDT. There is no detectable evidence of any trend.

However, by looking at the difference between the responsibility allocated to the dictator and to the computer it becomes clear that a large proportion of the passive dictators hold the dictator as far more responsible for the final outcome than the computer (see Figure 38).

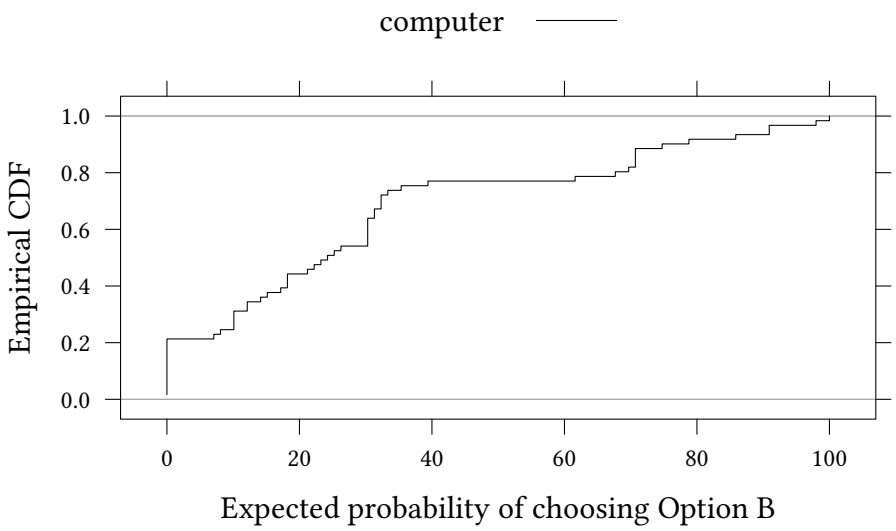

Figure 35: Expected human dictators' choice as seen by passive dictators
(Question 2 from Section A.1.2)

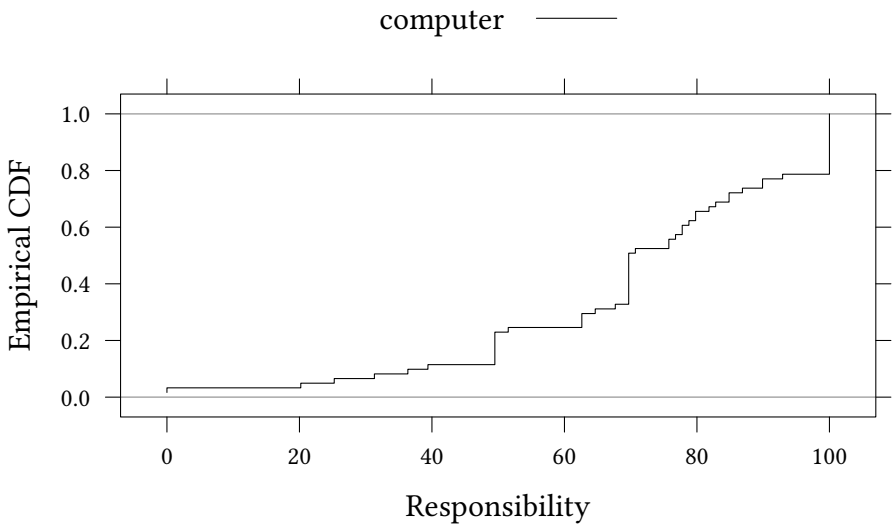

Figure 36: Responsibility allocated to the human dictator as seen by passive dictators
(Question 9 from Section A.1.2)

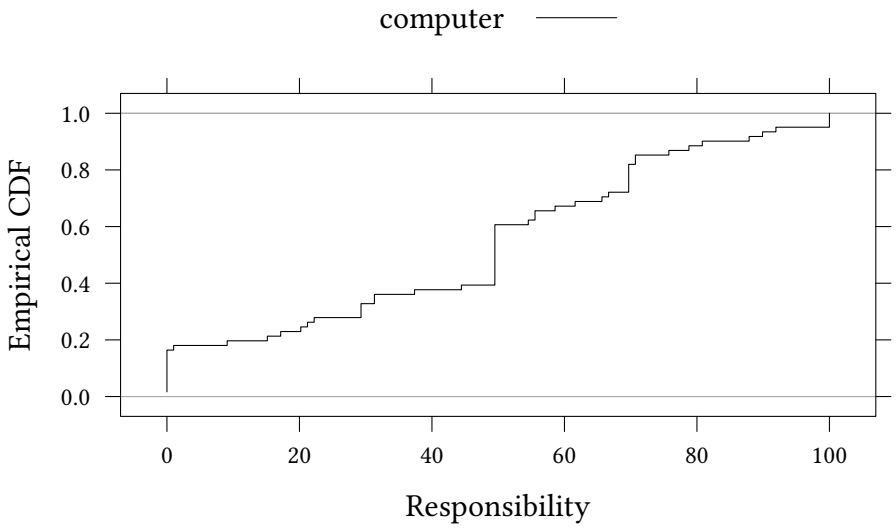

Figure 37: Responsibility allocated to the computer dictator as seen by passive dictators (Question 9 from Section A.1.2)

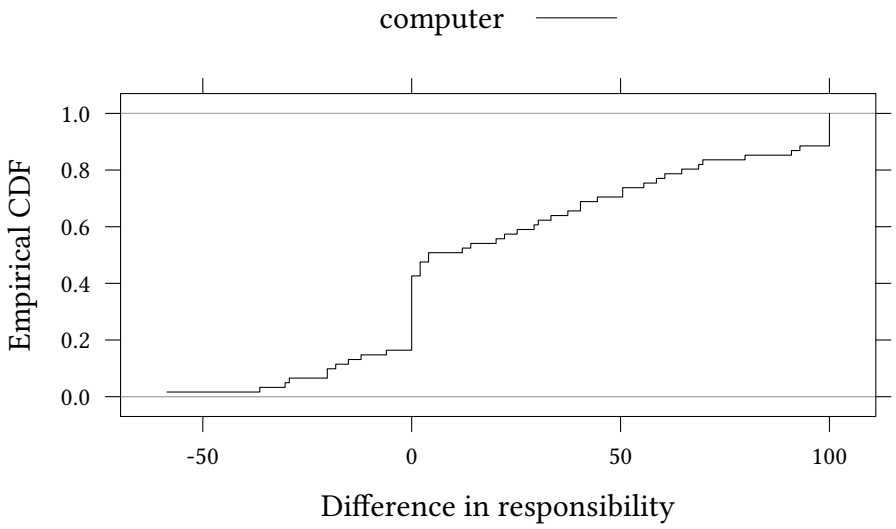

Figure 38: Difference in responsibility of the human dictator and the computer dictator as seen by passive dictators

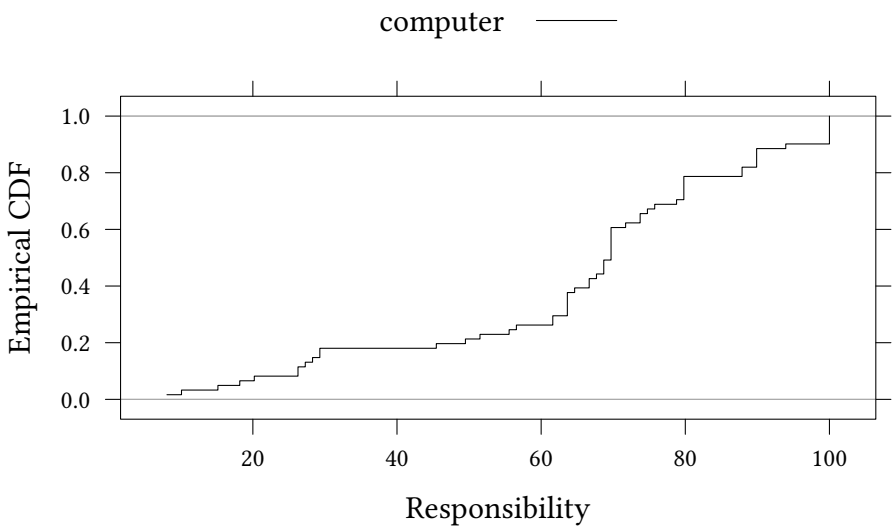

Figure 39: Allocated responsibility to the human dictator for the passive dictator as seen by passive dictators
(Question 7 from Section A.1.2)

A.6.4. Allocated Responsibility for the Other Dictator and Responder

The assigned responsibility was measured by a continuous scale from *"not responsible at all"* (0) to *"totally responsible"* (100). A large proportion of the passive dictators stated that they hold the dictator as very responsible for the payoff they receive as Figure 39 shows.

The result of the responsibility allocated by the passive dictators to the computer is shown in Figure 40. A large proportion of the passive dictators stated that they hold the computer also responsible for the the payoff they receive.

By looking at the difference between the responsibility allocated to the dictator and to the computer it becomes clear that a large proportion of the passive dictators hold the human dictator more responsible for their payoff than the computer as shown in Figure 41. However, the difference is not significant (p-value 0.0797).

The result for the responsibility allocated by the passive dictators for the responders' payoff to the dictator is shown in Figure 42.

The result for the responsibility allocated to the computer is shown in Figure 43. A large proportion of the passive dictators stated that they hold the dictator as very responsible and the computer as responsible for the final payoff the responder receives.

By looking at the difference between the responsibility for the payoff of the responder allocated to the dictator and to the computer by the passive dictator (see Figure 44) it becomes clear that a large proportion of the passive dictators hold the dictator more responsible for the payoff of the responder than the computer (p-value 0.0060).

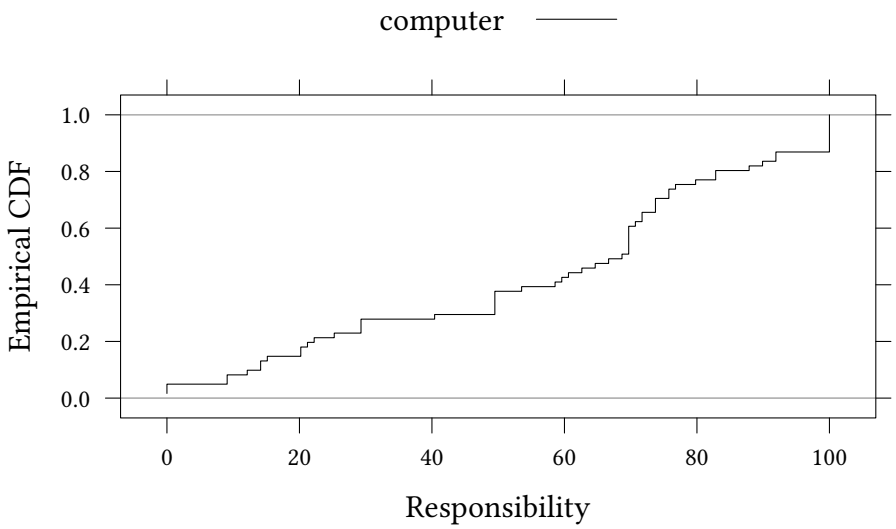

Figure 40: Allocated responsibility to the computer for the passive dictator as seen by passive dictators
(Question 7 from Section A.1.2)

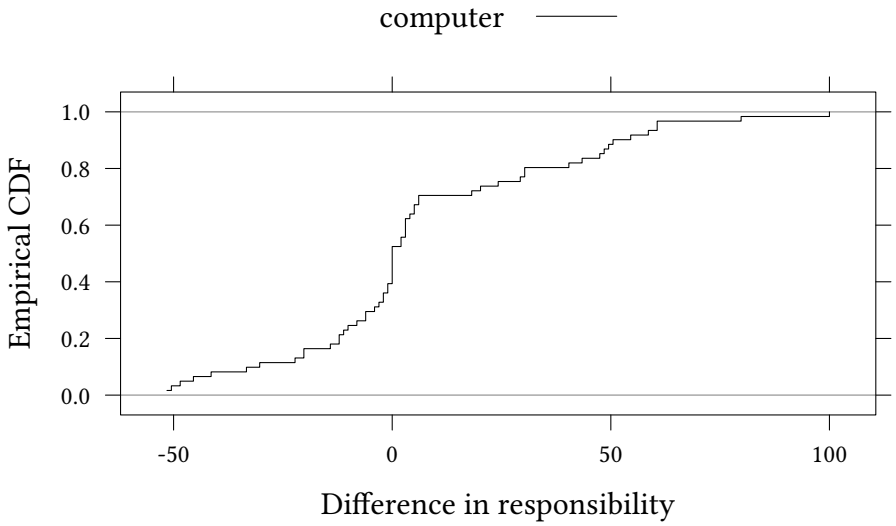

Figure 41: Difference in responsibility allocated to the human dictator and to the computer dictator for the passive dictator as seen by passive dictators

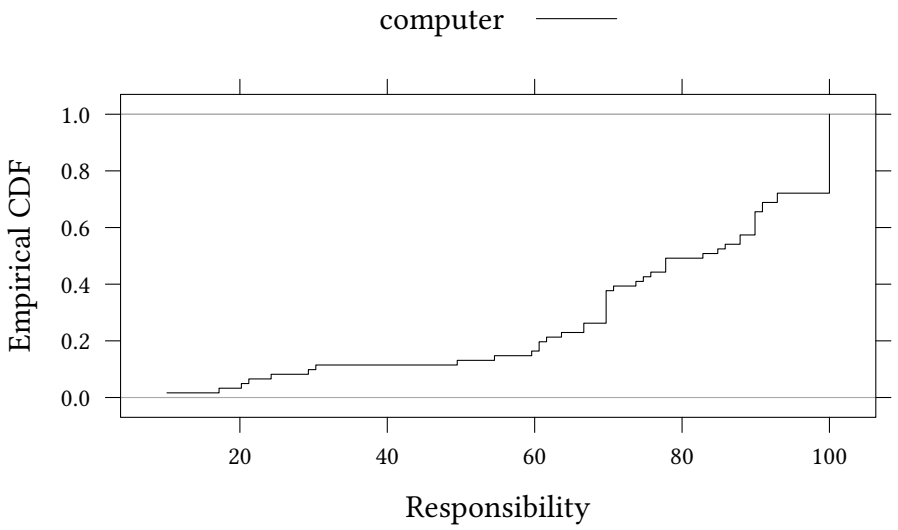

Figure 42: Allocated responsibility to the human dictator for the responder as seen by passive dictators
(Question 6 from Section A.1.2)

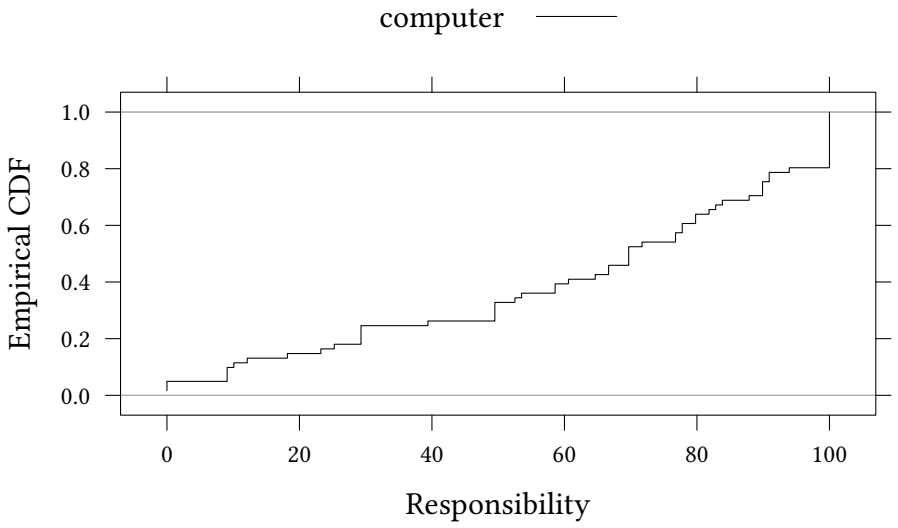

Figure 43: Allocated responsibility to the computer for the responder as seen by passive dictators
(Question 6 from Section A.1.2)

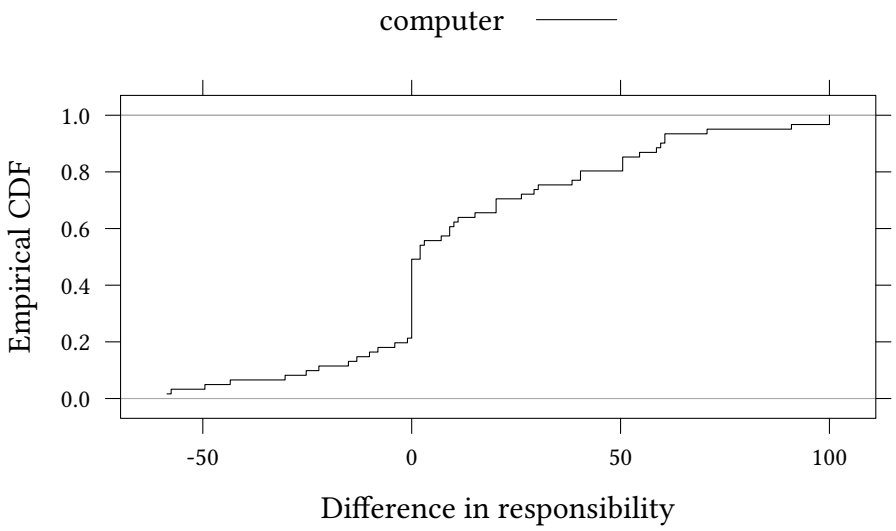

Figure 44: Difference in responsibility allocated to the human dictator and to the computer for the responder as seen by passive dictators

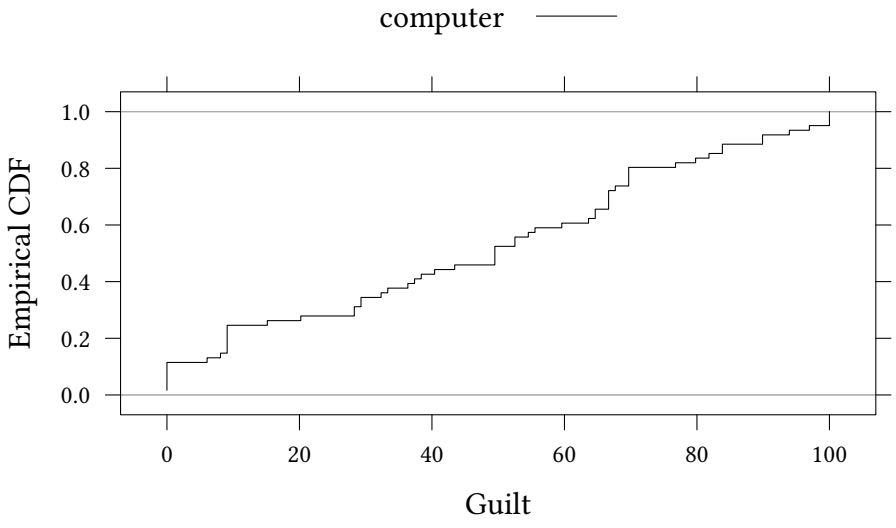

Figure 45: Expected guilt of the dictator as seen by passive dictators
(Question 8 from Section A.1.2)

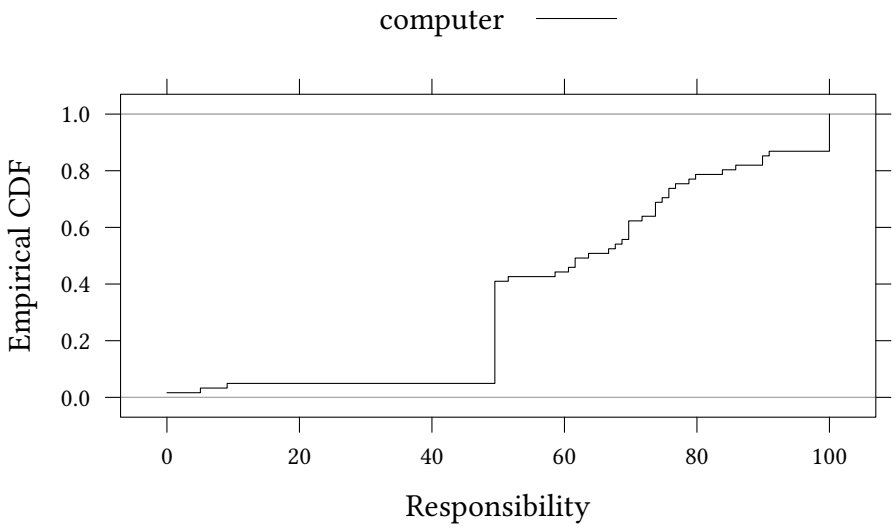

Figure 46: Expected human dictators' responsibility in the manipulation check as seen by passive dictators
(Question 4 from Section A.1.2)

A.6.5. Allocated Guilt

The perceived guilt was measured by a continuous scale from *"not guilty at all"* (0) to *"totally guilty"* (100). The allocated guilt to the dictator seems to be distributed more or less equally with no trend identifiable as shown in Figure 45.

A.7. Passive dictator: Manipulation Check

Passive dictators participating in the CDT were asked to state how responsible and guilty they think the dictator might perceive themselves for the outcome if, contrary to the game just played, the dictator would have to decide together with a another human instead of a computer. [35]

A.7.1. Manipulation Check: Allocated Responsibility to human Dictator(s) and Computer

How responsible the passive dictators expected the dictators to feel for the outcome in the manipulation check is shown in Figure 46.

For a comparison of the relative changes between the responsibility expected by the passive dictator to be perceived by the dictator(s) for the outcome in the hypothetical situation and in the actual experiment see Figure 47. A large proportion of the passive dictators expected the dictators to perceive themselves as less responsible if their counterpart is a human instead of a computer, however, the difference is not significant (p-value 0.0691).

[35] The wording of the manipulation check was the same as for responders.

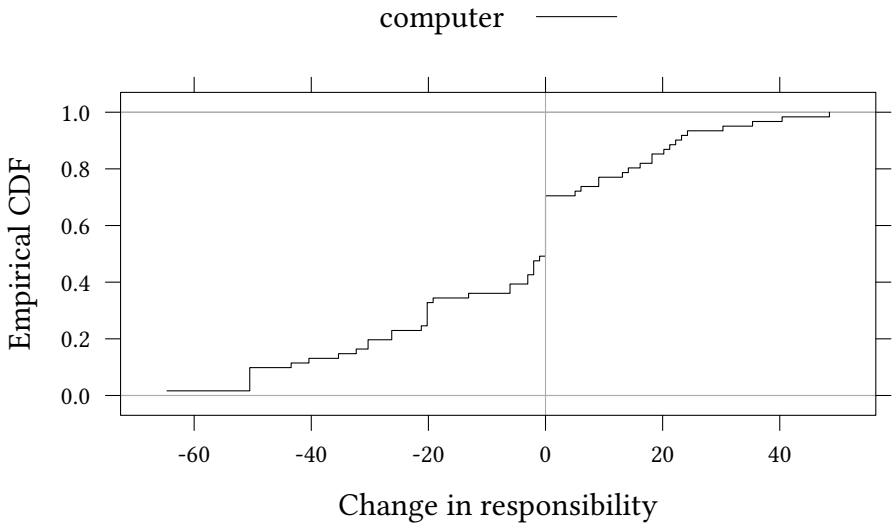

Figure 47: Change in allocated responsibility to the human dictator in the manipulation check as seen by passive dictators

The Figure shows the difference in the responsibility that the passive dictator expect the dictator to perceive for the decision between the hypothetical situation (described in Section A.3) and the actual experiment (as shown in Figure 38).

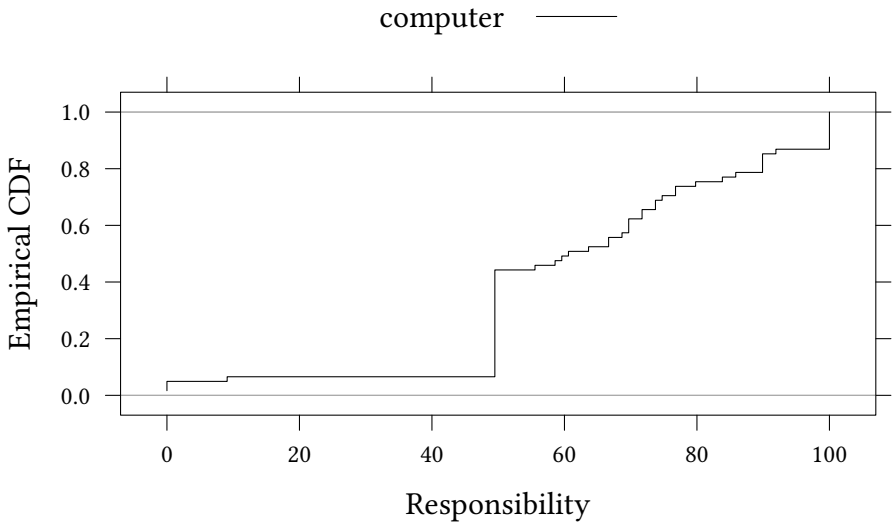

Figure 48: Expected responsibility allocted to the other human dictator in the manipulation check as seen by passive dictators

(Question 4 from Section A.1.2)

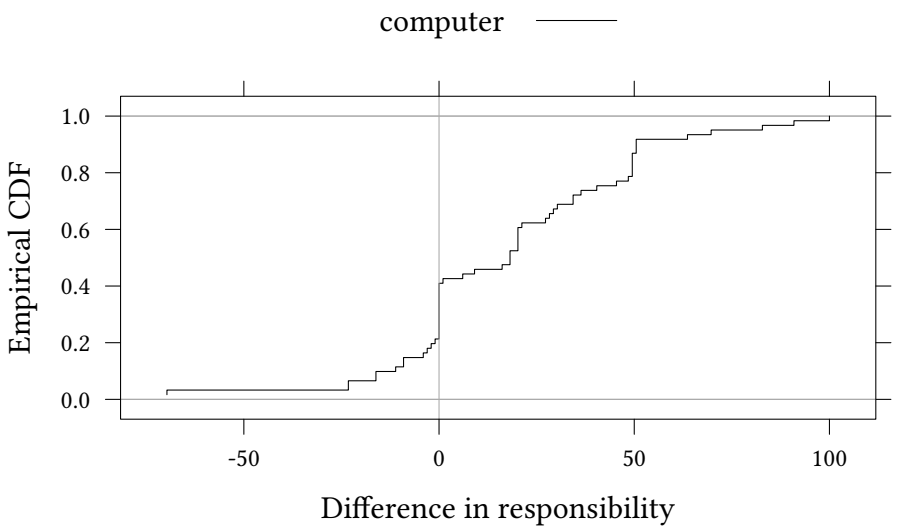

Figure 49: Difference in responsibility allocated to the human or computer dictator in the manipulation check as seen by passive dictators

The Figure shows the difference in the responsibility allocated by the passive dictator to the human dictator between the hypothetical situation (described in Section A.3) and the actual experiment (as shown in Figure 38).

The responsibility expected by the passive dictators to be perceived by the now human dictator for the final payoff in the manipulation check is shown in Figure 48. Passive dictators perceived the human dictator in the CDT to be very responsible for the final payoff.

For a comparison of the relative changes in the responsiblility allocated between the human dictator(s) in the hypothetical situation and the computer in the actual experiment see Figure 49. It can be seen clearly, that passive dictators perceived a human dictator to be more responsible for the final decision than a computer (p-value 0.0002).

A.7.2. Manipulation Check: Responsibility for the Passive Dictator

The perceived responsibility was measured from *"Same level of responsibility as in the experiment before"*, *"More responsible as in the experiment before"* and *"Less responsible as in the experiment before"*. Hypothesis 1.ii would suggest that dictators in the CDT who would share their decision with a human instead of a computer would feel less responsible than before for the payoff of the other dictator as Figure 50 shows. This was also expected by the passive dictators (p-value from a binomial test 0.0003).

A.7.3. Manipulation Check: Responsibility for the Responder

The perceived responsibility was measured from *"Same level of responsibility as in the experiment before"*, *"More responsible as in the experiment before"* and *"Less responsible as in the experiment before"*. Hypothesis 1.ii suggests that dictators in the CDT who share their decision with a human instead of a computer feel less responsible than before for the payoff of

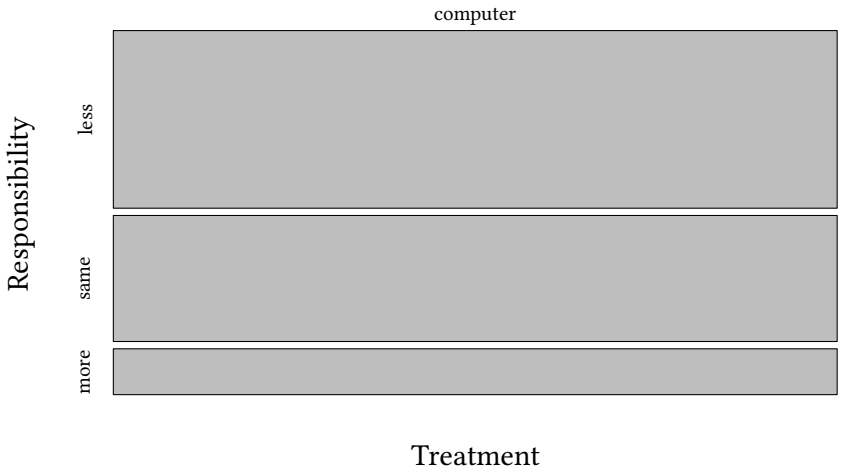

Figure 50: Change in responsibility perceived by the dictator for the passive dictator in the manipulation check as seen by passive dictators
(Question 2 from Section A.1.2)

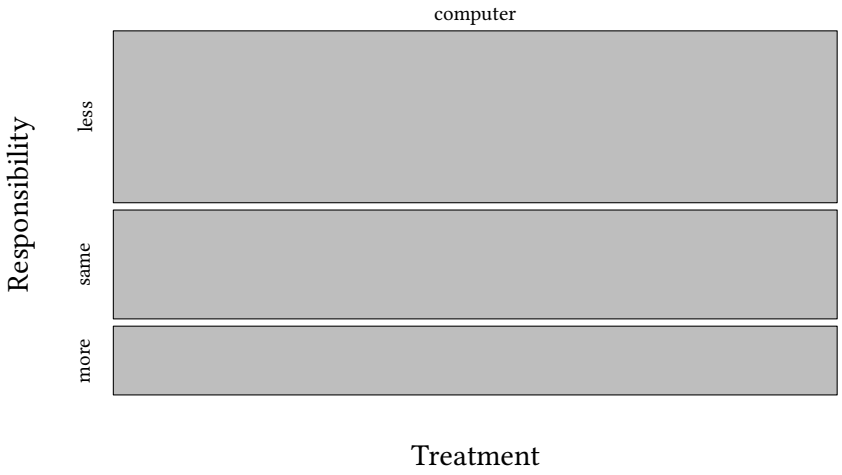

Figure 51: Change in responsibility perceived by the dictator for the responder in the manipulation check as seen by passive dictators
((Question 2 from Section A.1.2)

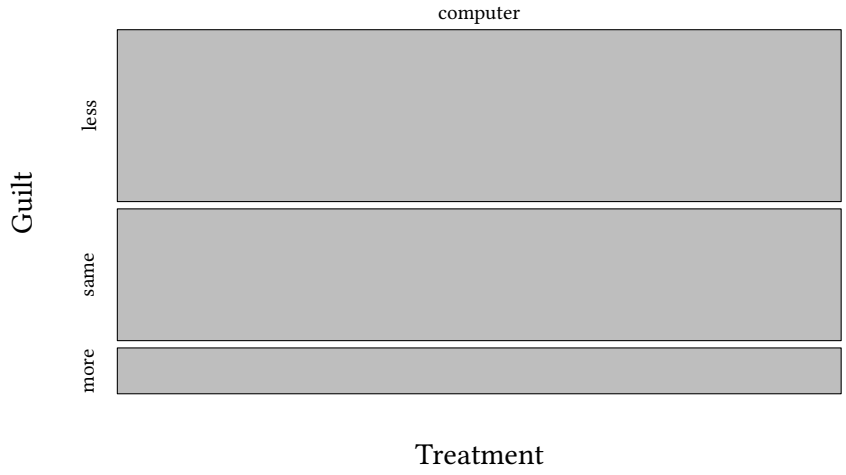

Figure 52: Change in guilt perceived by the dictator in the manipulation check as seen by passive dictators
(Question 3 from Section A.1.2)

the responder. This was also expected by the passive dictators as Figure 51 shows (p-value from a binomial test 0.0079).

A.7.4. Manipulation Check: Perceived Guilt

The perceived guilt was measured as *"Same level of guilt as in the experiment before"*, *"More guilt as in the experiment before"* and *"Less guilt as in the experiment before"*. Details are shown in Figure 52. In line with Hypothesis 2.ii dictators in the CDT were expected to feel less guilty once they can share the burden of their choice with a human as Figure 52 shows. This was also expected by the passive dictators (p-value from a binomial test 0.0005).