Ignorance, Intention and Stochastic Outcomes

Jana Friedrichsen (FU Berlin, HU Berlin, DIW and WZB)
Katharina Momsen (University of Innsbruck)
Stefano Piasenti (HU Berlin and DIW)

Discussion Paper No. 330

June 21, 2022

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Georg Weizsäcker, Ph.D., Humboldt University Berlin, 10117 Berlin, Germany
info@rationality-and-competition.de

Ignorance, Intention and Stochastic Outcomes*

Jana Friedrichsen† Katharina Momsen‡ Stefano Piasenti§

June 2022

Abstract

In sequential interactions, both the agent's intention and the outcome of his choice may influence the principal's action. While outcomes are typically observable, intentions are more likely to be hidden, leaving potential wiggle room for the principal when deciding on a reciprocating action. We employ a controlled experiment to investigate how intentions and outcome affect the principal's actions and whether principals use hidden information as an excuse to behave more selfishly. We find that principals react mainly to the intention of the agent. When intentions are not revealed by default, principals tend to select into information based on their inclination to behave more prosocially. While information avoidance is frequent and selfishness is higher with hidden information, we do not find evidence of a strategic exploitation of moral wiggle room.

JEL Classification: D91, C91
Keywords: information avoidance, dictator game, moral wiggle room, intentions, reciprocity

*This study was preregistered in the AEA RCT registry with ID AEARCTR-0005368 and was approved by the Board for Ethical Questions in Science of the University of Innsbruck (certificate 02/2020). We thank the participants in seminars at Lancaster University, HU Berlin, and the University of Potsdam as well as at the Annual Meeting of the Social Science Committee of the Verein für Socialpolitik 2020, the ESA Global Meeting 2020, and the 2020 PhD Conference in Behavioural Science and Policy for valuable comments. Financial support by Deutsche Forschungsgemeinschaft through CRC-TRR 190 (project number 280092119) and from the Leibniz Association through SAW-2015-DIW-4 is gratefully acknowledged.

†Freie Universität Berlin, WZB Berlin Social Science Center, Humboldt-Universität zu Berlin, DIW Berlin, and CESifo

‡University of Innsbruck (corresponding author, Universitätsstraße 15, 6020 Innsbruck, Austria, katharina.momsen@uibk.ac.at)

§Humboldt-Universität zu Berlin and DIW Berlin

1 Introduction

For a long time, classical economic theory centered around the benchmark of the *homo economicus*, a perfectly rational selfish individual. More recently, a large literature shows that individuals do not just care about their own payoffs but exhibit other-regarding preferences. However, when given the opportunity to justify selfish behavior, they may make use of this moral wiggle room to maintain a positive self-image although they act egoistically (Grossman and Van Der Weele, 2017; Bénabou and Tirole, 2011). While pro-social behavior as well as potential deviations from pro-sociality are thoroughly investigated in dictator games (Dana et al., 2007), the focus of this paper lies on sequential interactions in a principal-agent setup. When both outcomes and intentions of the agent can be observed by the principal, Charness and Levine (2007) find that intentions matter more than outcomes for reciprocating actions. Inspired by the literature on strategic ignorance, we developed a new design related to Erkal et al. (2021) that limits the observability of the agent's action, thereby introducing a potential excuse for selfish behavior of the principal while also adding realism to the setting.

Examples of sequential interactions with potentially hidden information are widespread. Consider, for instance, the following standard employment situation that arises in firms: an employee can work hard to make a project succeed or he can be rather lazy. In both cases, other uncontrollable factors also determine the success of the project. Hence, even if the employee puts a lot of effort into the project, it may fail. Similarly, if he does not try hard, it may still be the case that he is lucky and the project succeeds. Observing only whether the project succeeded or not, the boss needs to determine the employee's bonus payment. To do so, she can either try to find out how much effort her employee exerted or she can determine the bonus payment without knowing if her employee worked hard. Remaining ignorant about the exerted effort, she may create some wiggle room to justify a lower bonus payment: in case of a negative outcome, she can attribute it to the potential lack of the employee's effort, while in case of a success, she could claim that luck, not effort was the main driving factor behind the outcome.

To investigate an abstract version of the above-described situation, we conducted a laboratory experiment. Subjects interact in pairs of two and sequentially make a decision that affects their own as well as their matched partner's payoff. The first player (the agent – he) can invest a large or a small share of his endowment into a joint project of which the payoff is split equally between both players. The probability with which the project succeeds or fails is influenced by the invested amount. Hence, the action of the first mover yields a stochastic outcome that alone does not reveal the chosen action. We alter the information that is available to the second player (the principal – she) in a between-subject design: she either observes both the first player's investment decision — his intention — and the payoff of the project — the outcome —

(FULLINFO) or only the outcome (HIDDENINFO). In the latter case she has the option to reveal the first player's choice after seeing the outcome and before deciding how to split a fixed endowment between herself and the first player.

We study if, and to what extent, the principal's behavior depends on the outcome of the agent's choice as well as his intention. We further investigate whether the principal remains ignorant about the agent's investment decision strategically and keeps a larger share of her endowment when the investment decision is hidden by default than if she is fully informed. In addition, we analyze whether the agent reacts to the difference in information available to the principal.

Our results confirm previous findings on the importance of intentions for behavior but do not fully support previous evidence on an outcome bias. When the principal observes both intention and outcome, she strongly rewards the agent's good intention by sharing a larger amount when the agent has chosen the expensive investment option (Result 1). While the principal shares a larger part of her endowment after a good than after a bad outcome, the difference is relatively small and not statistically significant (Result 2). This is in contrast to previous findings by Brownback and Kuhn (2019) where a significant outcome bias was observed. We conclude that the first player's intention has a larger effect on the second player's decision than the outcome of his investment decision – a result that is in line with the one in Charness and Levine (2007).[1]

In contrast to a large literature, we find no evidence in support of the hypothesis that the principal will exploit the moral wiggle room provided by the intention being hidden. In contrast, the principals' allocation decisions under full information are comparable to those from the treatment where information is hidden by default (Result 3). While many studies have found moral wiggling, there are other studies where behavior does not differ with the information condition (e.g. Bartling et al., 2013; Lind et al., 2019). We discuss possible explanations in Section 6.

We further observe that those subjects who self-select into being informed as principals in treatment HIDDENINFO tend to allocate more to the agent than principals in the FULLINFO treatment when confronted with a successful investment. While the effect is not statistically significant, we note that our sample is too small to rule out effects of relevant size as we are only well-powered to detect large to very large effects (Result 4).

Finally, even though subjects in the role of the agent choose the more expensive investment option less frequently when the outcome of their decision is initially hidden than when it is immediately observable, we cannot confirm that this difference is statistically significant (Result 5). This is consistent with the fact that we do not find

[1]This does not preclude the possibility that the outcome does have a small effect on the agent's sharing decision as our study is only well-powered to detect effects of medium size or larger. Details on our sample planning and the power of our analyses are provided in Appendix B.

evidence for principals exploiting the moral wiggle room provided by the treatment HIDDENINFO but we cannot rule out that there are small effects which would only be detected in a larger sample. Furthermore, we observe that principals are on average too optimistic and, for uninformed principals, the expected share of investing agents correlates positively with donations.

2 Related Literature

Our experimental design is inspired by Charness and Levine (2007) who also consider a game between two players where the second player can reciprocate the first player's action. The outcome of the first player's action depends both on his choice and on luck. In contrast to our paper, the second player can immediately observe the first player's action. Thus, there is no scope for motivated information avoidance. Charness and Levine (2007) find that the first player's intention has a large effect on the second player's decision, while the decision outcome only has a minor effect.[2] However, note that this finding is contested by the results in Pan and Xiao (2016) who find that in a gift exchange actual gifts create a larger impulse to reciprocate than intended gifts.

A growing body of literature investigates motivated reasoning in such sequential interactions. In a setting similar to Charness and Levine (2007), Erkal et al. (2021) investigate if the second player holds biased beliefs regarding the first player's action. Observing only the final payoff and ignorant about the decision maker's action, the second player tends to attribute good outcomes to luck and bad outcomes to intentional actions. Hence, decision makers receive too little credit for good outcomes and, relative to praise for success, they receive too much blame for failure. Erkal et al. (2021) argue that their experimental setting brings the literature on outcome bias closer to reality because, in many situations, the decision maker's choice remains concealed. Following their experimental setting, our design takes even one step closer to depicting realistic decision situations, as the principal has the option to reveal the agent's choice.

In a setting related to ours, Brownback and Kuhn (2019) investigate outcome bias in situations where the agent's action is always immediately observable and it is the outcome of his investment decision that may be hidden. Hence, their design constitutes the exact opposite to ours. Moreover, while principals in our experiment can reward agents by generously splitting their endowment, principals in Brownback and Kuhn (2019) have the possibility to punish agents. They find particularly strong evidence of outcome bias even if both outcome and investment decision are observable. If the outcome is hidden, principals are more responsive to effort.

[2] In a different context, Falk et al. (2008) study whether individuals respond to fairness in intentions or in outcomes. They find that second movers reciprocate first mover actions almost one to one if they can infer intentions, but they do much less so otherwise.

Slightly further away from our design is the study on blame by Gurdal et al. (2013). In their experiment, the agent can choose between a safe and a risky option that determine the principal's payoff. The authors find that principals are less generous towards the agent if their payoff had been higher in case the agent had chosen the other option. Hence, they blame the agent for an outcome for which he is not responsible. In our setting, agents choose between two risky options with the counterfactual and, depending on the treatment, the agent's choice initially unknown to the principal. Another difference is that, in our study, principals play a dictator game with the agent as recipient and no additional third party involved. Yet blame may also play an important role in our study: remaining ignorant and observing a bad outcome, a principal may blame the agent and therefore transfer only a small amount, even though the agent may have chosen the high cost investment.

While Gurdal et al. (2013) suggest that principals blame the agent, they do not explicitly study the underlying beliefs. Ging-Jehli et al. (2020) explicitly investigate the motivated reasoning that would correspond to such blaming in a two-player game. Specifically, they investigate whether subjects are strategically cynical with respect to another one's hidden action to justify more selfish behavior. However, they find no evidence in this regard: players are not strategically pessimistic about the other's kindness. In our experiment, a cynical first player would expect the second player to exploit moral wiggle room regarding the first player's investment decision to justify selfish behavior. In this case, he would be less willing to take the more expensive investment decision. Further, a cynical uninformed second player would expect the first player not to have invested and hence choose a low donation.

In a broader framework, our paper relates to the large literature on the role of other-regarding preferences and social context.[3] Still, selfish interests are an important driver of behavior, with existing experimental work highlighting the possible conflict that results from egoistic and social considerations. In situations that allow individuals to choose a selfish action while maintaining a good self-image or appearing "good" to potential observers, average behavior is less pro-social than in situations where such moral wiggle room does not exist. Originally identified and studied in the context of dictator games (Dana et al., 2007; Larson and Capra, 2009; Feiler, 2014; Grossman, 2014), strategic ignorance and the exploitation of moral wiggle room are also observed in trust games (Regner, 2018), donations to charity (Exley, 2016), and contributions to

[3]For instance, experimental studies find that subjects are willing to sacrifice own payoffs to increase social welfare as they value efficiency, that they reciprocate positively as well as negatively, and that they care about payoff inequality and payoffs to the least well-off (Charness and Rabin, 2002; Engelmann and Strobel, 2004; Fehr and Schmidt, 2006; Güth and Kocher, 2014). Further studies show that participants cooperate in prisoner's dilemmas and contribute to public goods inside and outside the laboratory beyond the selfishly rational benchmark (Andreoni and Miller, 1993; Ledyard, 1994; Henrich et al., 2001; Shang and Croson, 2009; Chaudhuri, 2011).

carbon offsets (Momsen and Ohndorf, 2020).[4] Interestingly, they are not only observed in situations characterized by *ex-ante* uncertainty about the recipient's payoff but also when *ex-post* information about the recipient's true needs is obscured (Kandul, 2016). Building on this literature, the aim of our paper is to investigate decisions in sequential two-player interactions when a potential excuse for not reciprocating pro-social behavior may be available.

3 Experimental Design and Procedures

The experiment is designed to investigate how decision makers who move second in a principal-agent setting take both the agent's intention and the outcome of his choice into account when choosing a reciprocating action. In a between-subjects design, we vary whether the principal observes the agent's decision or only the stochastic outcome of it.

Subjects interacted taking the roles of agents (player 1) and principals (player 2). They were randomly assigned to their roles at the beginning of a session and kept their roles throughout the experiment. Subjects interacted with a participant in the opposite role four times. For each interaction, pairs were formed anew following a perfect stranger matching protocol. Participants' identities remained anonymous throughout the entire experiment. There was no feedback about actions or payoffs between the four rounds. Within each round, each pair engaged in a sequential game. After the four rounds of interaction, risk preferences as well as beliefs about player 1's behavior were elicited. The experiment concluded with a questionnaire.

Agent: Player 1 made an investment decision with a stochastic outcome. He received an endowment of 30 points from which he could choose to invest a high amount of 25 points or a low amount of only 5 points.[5] If the investment succeeded, the return of the project was high (= 50 points); if the investment failed, the return was low (= 10 points). With a probability of 75%, the chances for success were higher if player 1 chose the high investment compared to a 25% success probability if the low investment was chosen. In either case, the return of the project was split equally between both players.

Principal: After player 1 made his decision, player 2 received a separate endowment of 30 points and played a dictator game with player 1. In this game, player 1 took the role of the recipient, while player 2 had to decide how to allocate her endowment between herself and player 1. Any integer amount between 0 and 25 points was possible

[4]For an overview of the literature on information avoidance see Golman et al. (2017).

[5]The decisions were labelled neutrally, i.e. the high investment decision was called "Investment X" while the low investment was referred to as "Investment Y" as shown in the instructions in Appendix D.

so that the second player had to keep at least 5 points to herself.[6]

Note that choosing "Investment X" over "Investment Y" costs player 1 20 points but only increases his payoff from the project by 10 points in expectation. Choosing the high investment, however, improves the expected payoff of player 2 in the first stage of the experiment and has the potential to increase player 1's payoff if player 2 reciprocates. Thus, in order to be willing to choose "Investment X", player 1 must be sufficiently confident that player 2 will reciprocate the generous action and compensate player 1 with a higher share of her endowment. Since, overall, the expected payoff from choosing "Investment X" and "Investment Y" is identical, efficiency considerations on the part of player 1 cannot affect behavior.

Round payoffs: In each round, the payoff of player 1 equaled her endowment of 30 points *minus* the investment cost (high or low) *plus* half the realized return of the project *plus* the donation she received from player 2 in the dictator game. The round payoff of player 2 equaled half the return of the project *plus* her endowment from the dictator game endowment *minus* the donation to player 1.

Beliefs and risk preferences: After the last round, we elicited subjects' beliefs about the investment decisions of player 1 as well as their risk preferences. Both tasks were incentivized and one was randomly determined to be payoff-relevant at the end of the experiment. For a measure of beliefs about player 1's choices, subjects had to guess the proportion of subjects in the role of the agent who chose the high cost investment in the first round. This question refers to behavior in round 1 only in order to at least partially mitigate the concern that beliefs are determined by experience made during the four rounds.[7] We use a choice-list design; a screenshot of the decision screen as well as a detailed description of the task is provided in Appendix D.

Treatment variation: In a between-subjects design, we vary whether or not information about the agent's investment choice was available to the principal. In both treatments, the principal observes the outcome of the agent's investment decision.

[6]Note that the principal's action space is independent of the agent's investment decision, as the possible amounts to give to player 1 do not depend on any payoff received from player 1's investment choice. In other words the principals have – regardless of the outcome of agent's investment decision – always the same budget that they can split between themselves and the agent. This design choice ensures that any change in the principal's action is not driven by a mechanical effect from the action space having changed.

Further, note that the principal could only give integer amounts to the agent. In the instructions (see Section D in the Appendix), we use an impossible even split of 12.5-12.5 as an example. We do so to give a concrete example to facilitate the comprehension while limiting potential anchoring affects.

[7]In the final data, we do not find any evidence for an effect of experience over the four rounds on beliefs for round 1.

In the treatment FULLINFO, we use the strategy method to elicit decisions *as if* the principal had full information about the agent's investment decision. Specifically, having observed the realized outcome, the principal specifies a donation for each possible investment decision of the agent — high or low. Which allocation decision is implemented depends on the agent's actual investment decision. This method also allows us to observe behavior at rarely reached nodes of the game, e.g. behavior following an unsuccessful high-cost investment decision.[8]

In the treatment HIDDENINFO, having observed the realized outcome, the principal takes only one decision, by default not knowing which investment the agent has chosen. While the principal does not observe the agent's investment decision upfront, she can click a button to reveal it at a payoff-irrelevant cost.[9] Yet, the principal can also refrain from revealing the agent's choice and decide solely based on the outcome of the hidden investment decision.[10] Willingly informed players are the only ones in the treatment with HIDDENINFO to ever learn about an investment decision of player 1.

Final payoffs: A subject's payoff from the experiment consisted of the payoff from one randomly selected round out of four rounds of interaction in pairs plus the payment from either the belief elicitation task or the task measuring risk preferences. This payoff in points was converted into euros with an exchange rate of 1 point = 0.2 Euros. In addition to the experimental payoff, each participant received a show-up fee of 5 Euro.

Procedures: We collected data in 12 experimental sessions conducted in the experimental economics laboratory at Technical University Berlin and in seven sessions conducted in the PLEx laboratory at the University of Potsdam in February 2020. We ran five sessions of FULLINFO and 14 sessions of HIDDENINFO with 18 to 22 participants each. The total data set comprises decisions from 374 subjects such that the realized number of participants falls below our preregistered target sample size.[11] The closure of the laboratories as part of the measures against COVID-19 prevented additional data

[8]Taking up concerns that subjects' decisions might differ when using the strategy method as compared to the direct response method, we discuss the potential effects of this design choice in the context of our results.

[9]Clicking the button was associated with a cost of 0.1 points. With an exchange rate of 10 points to 2 Euros, the cost of revealing information was equal to 2 Cents. Subjects were informed in the instructions that their final earnings in Euros would be rounded up to the next 10 Cents. Since only one round was payoff-relevant, clicking could not reduce their final payoff. We implemented this small, payoff-irrelevant cost to capture the fact that information on the agent's action is often available, yet it takes a negligible amount of effort to gather, which may be taken as an excuse to remain ignorant.

[10]Screenshots of Player 1's and Player 2's decision situations both in FULLINFO and HIDDENINFO are reported in Figure 9, Figure 10 and Figure 11, respectively, in the Appendix.

[11]Link to preregistration: `https://www.socialscienceregistry.org/trials/5368`

collection. We describe the implications for the power of our analyses in Section B in the Appendix.

The experiment was programmed in zTree (Fischbacher, 2007) and participants were recruited using ORSEE (Greiner, 2015). At the beginning of each session, subjects received detailed written instructions about the experiment. A translation of the original German instructions is included in Appendix D. The experiment only started once all participants had correctly answered a set of control questions. Sessions lasted approximately 60 minutes (including payment) and average payment was 15.20 €. On average, player 2 earned more (15.50 €) than player 1 (14.90 €).

4 Behavioral Predictions

In light of an important strand of the literature focusing on the role of intentions in a two-person trust game (see, for example, McCabe et al. 2003, Toussaert 2017), we investigate if the second player values good intentions. We hypothesize that, for identical outcomes, player 2 allocates a larger share of her endowment to the first player when he has chosen the costly investment option as opposed to an interaction with a first player who has chosen the cheap investment option.

Hypothesis 1. *Conditional on the realized outcome, average donations in treatment* FULLINFO *are higher in decisions where player 1 has chosen a high investment than in those with a low investment (Rewarding Intentions).*

Following Brownback and Kuhn (2019), we further analyze if the second player exhibits an outcome bias, i.e. if, given identical actions of the first player, she shares a larger part of her endowment with the first player if the investment was successful. The success of the investment decision does not influence the sum the second player can split as she receives an endowment for her distribution decision that is independent of player 1's behavior. But the outcome from player 1's investment affects player 2's payoff and could therefore also affect her willingness to give to player 1.[12] Given previous evidence, we formulate the following hypothesis:

Hypothesis 2. *Conditional on the investment decision, donations of player 2 in treatment* FULLINFO *are higher if the investment succeeds than when it fails (Outcome Bias).*

According to cognitive dissonance theory, agents suffer in situations with conflicting motives (Festinger, 1957). If the first player has chosen the expensive investment

[12]Instead of irrationally reacting to the outcome, which is suggested by the term "bias", giving more after a successful investment would also result if player 2's marginal utility from sharing increases in her own payoff or if she wanted to secure a certain minimum payment to herself from the experiment that restricted her donations to player 1 after a failure. We acknowledge this alternative interpretation but still use the term employed in the related literature.

option, decision makers may experience a conflict between maximizing their own monetary payoff by keeping a large share of their endowment and reciprocating the prosocial behavior of the first player thereby reducing their own monetary payoff. When the first player's investment is hidden, however, they have the possibility to circumvent the potential cognitive dissonance: they can choose to remain ignorant about the first player's decision and, thus, maintain a positive self-image while acting egoistically. Therefore, we hypothesize that the average share of the endowment which the second player keeps for herself is larger under hidden than under full information.

Hypothesis 3. *Conditional on the realized outcome, donations in treatment* HIDDENINFO *are lower on average than those in treatment* FULLINFO *(Exploitation of Moral Wiggle Room).*

Grossman and Van Der Weele (2017) and Kajackaite (2015) show both theoretically and empirically that less pro-social types sort into ignorance while highly prosocial types sort into being informed. In line with this literature, we investigate whether second players who reveal the first player's intention are, on average, more generous. Thus, we compare the average donations of exogenously informed players in the full information treatment with willingly informed players in the treatment with hidden information, conditioning on investment decision and outcome. We also investigate whether second players who avoid information on the first player's investment decision behave more selfishly. To do so, we compare the average donations of players in the full information treatment to the donations made by willingly uninformed players in the hidden information treatment.

Hypothesis 4. *Average donations by informed (uninformed) player 2 in treatment* HIDDENINFO *are higher (lower) than those by exogenously informed players 2 in treatment* FULLINFO.

We also investigate if the first player's investment decision depends on the treatment. If the first player expects the second player to be less generous in the dictator game under initially hidden information about his investment, he might refrain from incurring the higher investment costs to improve the second player's payoff from the first part of the game. Hence, we expect to observe fewer choices of the high investment option in HIDDENINFO.[13]

Hypothesis 5. *Player 1 chooses the high investment less frequently in* HIDDENINFO *than in* FULLINFO.

[13]There is another potential explanation for this behavior: If player 2 remains uninformed, she will never know if the observed outcome results from a high or a low investment. Hence, a player 1 with social image concerns who expects player 2 to remain uninformed no longer feels social pressure to make a high-cost investment decision, see also Gueth et al. (1996).

5 Analysis and results

Our sample consists of 374 subjects.[14] 100 subjects took part in treatment FULLINFO and 274 in HIDDENINFO. Half of the subjects in either treatment made decisions in the role of player 1 and the other half in the role of player 2. This yields 187 observations for player 1 and 187 observations for player 2, with 50 for each type of player in treatment FULLINFO and 137 in treatment HIDDENINFO.

Each player made decisions in four rounds. Since decisions taken by the same individual in subsequent rounds are likely to be correlated, they cannot be treated as independent. Whenever we use non-parametric tests, we therefore compute subject-level averages for the considered decision situations. When we use regression analyses, we rely on panel methods to take the repeated observations into account. For the analyses with subject-level averages, we will state explicitly how we construct the respective averages for each analysis. Depending on the hypothesis tested, we condition on the investment outcome, the investment decision of player 1, or the information that player 2 has about this decision. Table 4 in the Appendix provides information on the number of subjects in the role of player 2 who took decisions for low and high outcomes, as well as low and high investments of player 1, respectively. Of the 50 subjects taking the role of player 2 in treatment FULLINFO, 48 (96%) faced at least one failed investment and 40 (80%) faced a successful investment at least once. As we employ the strategy method in the FULLINFO treatment, each player 2 takes a decision both for the low and the high cost investment decisions for the observed outcome.

5.1 Donation decision of player 2

Subjects in the role of player 2 decide in a standard dictator game how much of their 30-point endowment to give to player 1. By design, they always keep at least 5 points of their endowment. Hence, we focus our analysis on the 25 points that they can split between themselves and player 1. To account for potential correlation in the repeated decisions of subjects over the four rounds, we first present results from non-parametric hypothesis tests and complement these with regression analyses on the full panel of individual decisions.[15]

First, we analyze behavior in the treatment FULLINFO alone to investigate Hypothesis 1, which states that subjects in the role of player 2 will be more generous when player 1 has chosen the high investment than when she has chosen the low investment. Our sample contains decisions from 48 subjects where the investment of player

[14] Descriptive statistics for our sample are given in Table 3 in the Appendix.

[15] For the non-parametric tests, we collapse the data set to individual-level averages for the respective decision situation so that we have only one observation per individual per condition. This is necessary because the non-parametric tests would otherwise be overpowered as they do not allow us to correct for the potential correlation of decisions within the same individual.

1 has failed and from 40 subjects where the investment of player 1 has succeeded. The average donation after a failed investment is 5.46 points conditional on a high cost investment and only 1.93 points conditional on the low cost one. The difference in donations is highly significant (two-sided Wilcoxon signed-rank test, $z = 4.43$, $p < 0.001$) and amounts to player 2 reimbursing player 1 for 17.7% of the cost difference. Similarly, the average donation following a successful investment amounts to 6.85 points conditional on a high cost investment, but only 2.49 points conditional on a low cost one. Donation behavior is again highly significantly different (two-sided Wilcoxon signed-rank test, $z = 4.45$, $p < 0.001$) and the difference corresponds to 21.8% of the cost difference (see Table 5 in the Appendix). This result is confirmed by regression analysis that allows us to exploit the panel structure of our data and to control for potential time trends. Using a random-effects model, we regress the second mover's donation on the first mover's investment decision, its outcome and the interaction of these two factors, where "Investment" and "Success" denote dummy variables taking a value of 1 when player 1 opted for the high cost investment and the investment happened to be successful, respectively.[16] As can be seen in the upper panel of the Table 1, the regression results are fully in line with those from the non-parametric tests with the donation of player 2 being significantly higher for a high investment independent of whether it succeeds or fails.

Thus, our data support Hypothesis 1 that informed players reward intentions by responding to high cost investments with an increase in their donation.[17]

Result 1. *Conditional on the realized outcome, donations in treatment* FULLINFO *are on average significantly higher in decisions where player 1 chose a high investment than in those with a low investment. This holds both for successful and for unsuccessful investments.*

Next, we turn to Hypothesis 2, which states that, conditional on the investment decision, donations of player 2 in treatment FULLINFO are higher when the investment succeeds than when it fails. Out of the 50 subjects in the role of player 2, 48 made at least one decision for a failed investment and 40 made at least one decision for a successful investment (see Table 5 in Appendix A). Donations are on average higher after a successful investment than after a failed investment, but the difference is relatively small with an average increase of about 0.5 points in case of a low investment and about 1.4 points in case of a high investment.

[16]Since the Hausman test is only valid under the homoscedasticity assumption, we use both the Mundlak approach (Mundlak, 1978) and a test of overidentifying restrictions (see e.g. Schaffer and Stillman 2006) to choose between fixed-effect and random effect specifications. Both tests do not reject the hypothesis that differences between coefficients from fixed and random effects are unsystematic such that we employ a random effects specification.

[17]We repeat the analysis behind Table 1 while additionally controlling for whether a player has been exposed to a successful or unsuccessful investment in t-1. Results are displayed in Table 13 in the Appendix, confirming Result 1.

Table 1: Marginal effects of player 1 investment outcome and investment decision on player 2 donations

	Full Information	Willingly Informed Hidden Information
Investment		
Success=0	3.801***	9.238**
	(0.751)	(3.442)
Success=1	3.713***	5.707***
	(0.697)	(1.051)
Success		
Investment=0	0.690	1.461
	(0.564)	(0.944)
Investment=1	0.602	-2.069
	(0.594)	(3.857)
Observations	400	148
No. of Subjects	50	63

Notes: Dependent variable is player 2 donation. *Investment* and *Success* are dummies for the investment having been high and successful respectively. Output from random-effects regressions (marginal effects) as detailed in Table 8. Column 1 reports the effect of *Investment* conditional on *Success* being 0 or 1 and of *Success* conditional on *Investment* in treatment FULLINFO. Column 2 reports the effects from the same exercise using observations only from willingly informed players in treatment HIDDENINFO. Cluster-robust standard errors in parentheses (clustered on subject-level). The full estimation results are shown in Table 8 in the Appendix.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

To formally test Hypothesis 2, we restrict our attention to those 38 subjects who made at least one decision for a successful investment and also at least one decision for a failed investment and use the signed-rank test.[18] In case of a high cost investment, we find that the donation of player 2 is on average 7.13 (SD=5.37) after a successful investment and on average 5.99 (SD=5.48) after a failure. Thereby donations are significantly more generous after a success than a failure for high cost investments (two-sided Wilcoxon signed-rank test, z = -2.497, p = 0.013). However, conditional on player 1 having chosen the low cost investment, we cannot reject equality of donations for the two possible outcomes (average donations are 2.62 (SD=4.44) after success and 2.16 (SD=3.86) after a failure; two-sided Wilcoxon signed-rank test, z = -1.317, p = 0.188).

However, there are two possible concerns regarding the non-parametric test here: First, the signed-rank test excludes observations from 12 subjects who made decisions for only one of the two possible investment outcome which might introduce a

[18]Note that of the 40 subjects that faced at least one successful investment, two never faced an unsuccessful investment. Hence, we consider the remaining 38 subjects who made decisions for each investment outcome.

bias. Second, subjects make four subsequent decisions which we average for the non-parametric test. This averaging may disguise a time trend that might be problematic if a) first movers become less likely to choose a high investment over time such that second movers are more likely to see a low outcome in later rounds and b) second movers become less generous in later rounds.

Therefore, we again complement the non-parametric analysis with a regression-based investigation.The marginal effects of the regression reported in Table 1 (column 1) suggest that neither conditioning on high nor on low cost investment decisions do second movers reward successful investments more than unsuccessful ones. We conclude that our subjects do not exhibit an outcome bias, contrary to Hypothesis 2.[19]

Result 2. *Both conditional on a low and high investment, the donations of player 2 do not differ significantly with the investment outcome.*

Note that both Result 1 and Result 2 are robust to controlling for the subjects' beliefs about the share of first movers choosing the high investment.

Before investigating our third hypothesis, let us take a step back and address two potential concerns that may arise in our analysis of Hypotheses 1 and 2: First, we employ the strategy method in the treatment FULLINFO for the second mover's donation decisions. While this method enables us to specify the effect of intentions on donations *within* subject and thus also allows us to economize on the number of observations we needed to collect, it may lead to behavior different from what we would find using a direct response method. To investigate if our result of decision makers rewarding good intentions merely constitutes an artefact of the strategy method, we consider the donation decisions of *willingly* informed subjects in treatment HIDDENINFO which uses the direct response method. The marginal effects presented in the upper part of Column 2 in Table 1 show that also willingly informed players in treatment HIDDENINFO reward good intentions. Furthermore, as depicted in the lower part of Column 2 in the same table, we do not find evidence that willingly informed subjects exhibit an outcome bias. Hence, the results are absolutely comparable to those from the strategy method in FULLINFO and we are confident that the use of the strategy method in treatment FULLINFO does not drive our results on rewarding good intentions and the absence of an outcome bias. We would like to further note that the evidence on differences in results between direct response and strategy method as discussed in Brandts and Charness (2011) would even suggest that subjects should more strongly reciprocate the intention of player 1 in a direct response design so that this particular design choice would work against us.

[19]We repeat the analysis depicted in Table 1 by additionally controlling for whether a player has been exposed to a successful or unsuccessful investment in t-1. Results are displayed in Table 13 in the Appendix and confirm the absence of an outcome bias.

Second, one may argue that, in addition to or instead of reciprocity considerations, Result 1 is driven by subjects being inequality averse: A simple desire to reduce inequality in final payoffs would lead to lower transfers conditional on a low cost investment – an effect pointing into the same direction as the desire to reward good intentions. To investigate if, in addition to rewarding good intentions, inequality aversion determines transfers, we use additional questionnaire data from the Berlin subsample and classify second movers into two categories based on two items that we expect to correlate with their inequality aversion. We perform this analysis only on the Berlin subsample because we do not have detailed questionnaire data from the Potsdam one. The variable *"Inequality"* captures participants' opinion on the statement: "Politics should equalize differences between large and small incomes". The variable *"Unemployment"* contains participants' opinion on the statement: "Unemployment must be tackled, even if that means high public debt." In both cases, we expect higher agreement from subjects who are more inequality averse.[20] Table 14 in the Appendix shows that even players 2 who (strongly) disagree with the two above-mentioned statements – and hence are supposedly not inequality averse – donate significantly more to player 1 when they chose the costlier action than when they chose the cheaper one. This suggests that rewarding intentions is a major factor driving Result 1 even if inequality aversion might additionally affect donations.[21]

We now turn to the third question, whether individuals exploit the moral wiggle room created by the first player's decision being hidden. We expect that subjects in treatment HIDDENINFO avoid learning about player 1's investment decision so as to justify on average lower donations that do not reward player 1's intentions. We investigate the corresponding Hypothesis 3 by comparing donations between the treatments FULLINFO and HIDDENINFO. We are interested in the aggregate effect, including also the possibility that subjects inform themselves before making their donation decision and, therefore, include both informed and uninformed players in the analysis.

Donations in FULLINFO are elicited for both high and low investments using the strategy method for a given outcome but not so in HIDDENINFO. This implies that uninformed participants in HIDDENINFO who face a given outcome should expect a high and a low cost investment with a certain probability, while participants in FULLINFO will not factor in the probabilities of the respective situation being payoff-relevant. To make the data from both treatments comparable to each other, we compute aver-

[20] We are aware of the fact that attitudes towards redistribution may also depend on beliefs about the determinants of income (see e.g. Alesina and Angeletos, 2005). Even though the two questionnaire items cannot take these beliefs into account, they might still function as a proxy for an overall attitude towards inequality in payoffs.

[21] We also check whether second movers transferred amounts that equalized payoffs between both parties. This appears not to be the case. The distributions of transfers as plotted in Figures 4 - 7 in the Appendix reveal no large spikes in transfers at those levels that equalize payoffs – especially not after a high investment and a low outcome.

age donations in the treatment FULLINFO using the mean empirical frequency of high and low cost investments conditional on the outcome being low or high, respectively, from the treatment HIDDENINFO.[22] The imputed average donation from treatment FULLINFO and the average donation observed in treatment HIDDENINFO, by design, incorporate the same distributions of high and low cost investments conditional on either investment outcome and allow us to compare donations conditional on outcomes alone across treatments.

We find that average donations tend to be *higher* in treatment HIDDENINFO than in FULLINFO for both low and high outcomes but the raw differences are not statistically different from zero for either of the two possible outcomes (two-sided Wilcoxon rank-sum tests: $z = -0.941$, $p = 0.35$ conditional on the low outcome and $z = -0.733$, $p = 0.46$ conditional on the high outcome).[23] Again we complement the non-parametric analysis of the averaged data with a regression analysis, which does not provide evidence in favor of Hypothesis 3 either.[24] If anything, donations are higher on average in the treatment with hidden information.[25]

Result 3. *We find no evidence that donations in treatment* HIDDENINFO *are lower on average than donations in* FULLINFO.

This result is surprising at first given the evidence on moral wiggling in other contexts and because the evidence on fully informed players strongly rewarding intentions (see Result 1) indicates that there would be something to gain from wiggling. We discuss potential explanations for this in the concluding discussion.

Next, we analyze the relation between the decision to become informed and donation behavior. We note that the vast majority of decisions were made while uninformed. If a *low* outcome was observed, player 2 chose to become informed about player 1's intention in only 25% of the decisions, i.e. in about one out of the four decisions that subjects made during the experiment. If the outcome was *high*, player 2 revealed the information in about 30% of the decisions on average.[26] We find no evidence that the information choices differ significantly with the observed investment outcome and neither do they become significantly less frequent over time.[27]

We then compare average donations of exogenously informed players 2 in FULLINFO with willingly informed players in HIDDENINFO. We perform the analysis for all pos-

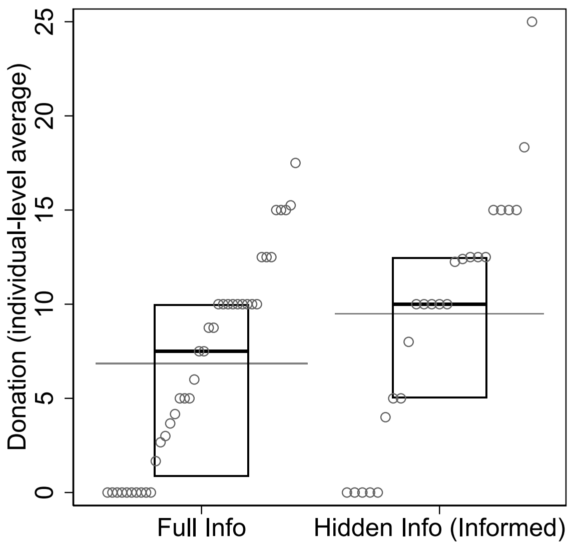
[22]These frequencies are summarizes in Table 6 in the Appendix.

[23]The respective descriptive data is collected in Table 7 in the Appendix.

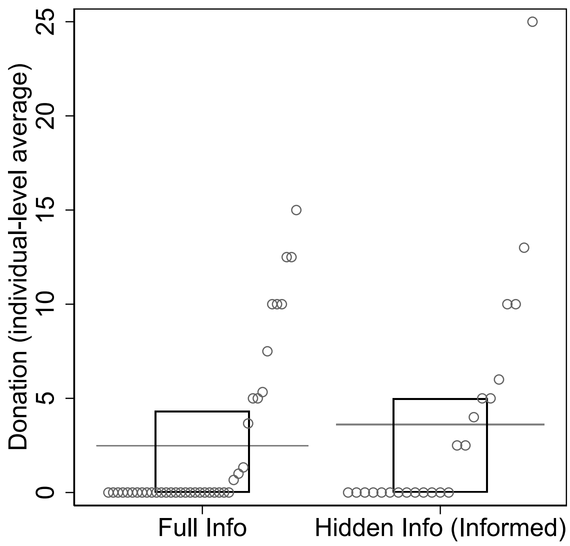
[24]The regression output is shown in Table 9 in the Appendix. We compute marginal effects for the variables of interest to confirm that the total effect of outcome on donations is not significantly different from zero in either treatment but intentions are rewarded less on average in HIDDENINFO than in FULLINFO (see Table 10 in the Appendix).

[25]When controlling for the subjects' beliefs on the share of first movers choosing the high investment, we find that donations in the FULLINFO treatment are significantly *lower* than under HIDDENINFO.

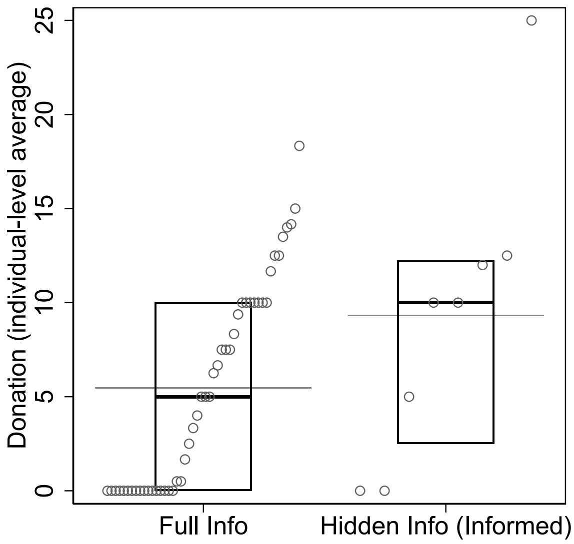
[26]The low share of subjects revealing information may be driven by the negligible revelation costs. However, the findings in Momsen and Ohndorf (2021) suggest that while negligible revelation costs decrease the share of subjects revealing information, they have little to no impact on the subject's actual



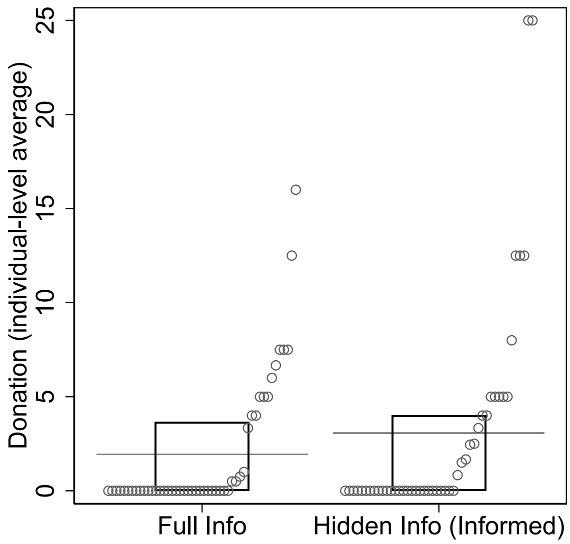
(a) High investment succeeds. N = 40 in FULLINFO and N = 25 in HIDDENINFO.



(b) Low investment succeeds. N = 40 in FULLINFO and N = 23 in HIDDENINFO.



(c) High investment fails. N = 48 in FULLINFO and N = 8 in HIDDENINFO.



(d) Low investment fails. N = 48 in FULLINFO and N = 46 in HIDDENINFO.

Figure 1: Player 2 donation after successful and unsuccessful investments split up by whether or not the investement succeeded; comparison between treatment FULLINFO and for informed players 2 in treatment HIDDENINFO. Each circle corresponds to one subject-level average. The boxes mark the interquartile range with a bold line at the median. The wide lines indicate the means. N states the number of subjects included.

sible constellations of investment and outcome. Figures 1a and 1b show that willfully informed players 2 in treatment HIDDENINFO tend to give more after a high outcome than those who are informed by default in FULLINFO, independent of the investment decision of player 1. Yet the differences are not statistically significant (two-sided Wilcoxon rank-sum test on individual-level averages in case of a high investment: z = 1.652, p = 0.099; in case of a low investment: z = 0.687, p = 0.492).

A similar picture obtains after a failed investment. Figures 1c and 1d show that willfully informed players 2 in treatment HIDDENINFO tend to give more after a failed investment than those who are informed by default in FULLINFO. Yet again, the differences are not statistically significant (two-sided Wilcoxon rank-sum test in case of a high investment: z = 1.281, p = 0.200; in case of a low investment: z = 0.666, p = 0.505). Regression analysis using all decisions instead of averages confirms the result (see Table 12, column 1, in the Appendix).

To compare donations of uninformed players to those from FULLINFO, we compute a weighted average of player 2 donations in FULLINFO that uses the frequencies of investments in HIDDENINFO. Doing so allows us to infer how players from FULLINFO would behave on average for investments comparable to those in HIDDENINFO and to analyze how decisions vary with treatment if the only difference was the information about player 1's intention.[28]

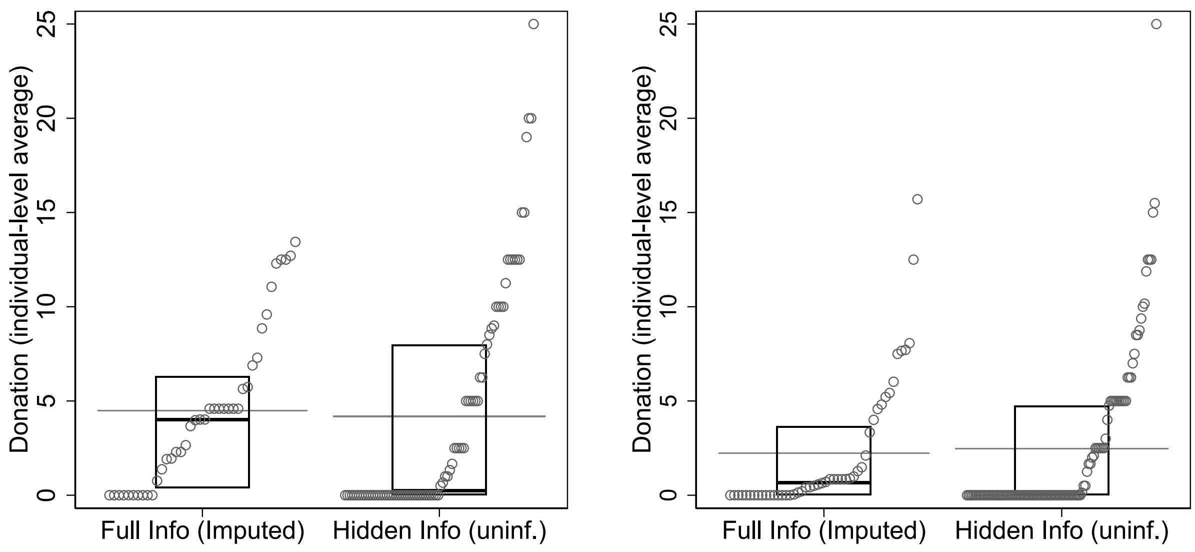
Figure 2a shows that when observing a high outcome, i.e. a successful investment, willfully uninformed players 2 in treatment HIDDENINFO choose slightly lower donations than what we would expect from players 2 in FULLINFO conditioning only on the investment outcome. The average donation for the willfully ignorant players is 4.19 points compared to 4.49 points for the latter. We cannot reject the null hypothesis that donations are equal (two-sided Wilcoxon rank-sum test, z = -1.591, p = 0.131). The difference goes into the opposite direction after an unsuccessful investment. Willfully uninformed players 2 in treatment HIDDENINFO give an average of 2.48 points after a low outcome, while informed players would give an average of 2.23 points after a low outcome (see Figure 2b). Again, we cannot reject the null hypothesis that donations are equal (Wilcoxon rank-sum test, z = -1.662, p = 0.099). Regression analysis confirms the result (see Table 12, column 2 in the Appendix).

To summarize, the differences in donation behavior are consistent with the idea that subjects who choose to learn about player 1's investment give more and reward inten-

choices. The presence of revelation costs only makes our null results on Hypothesis 3 and Hypothesis 4 stronger.

[27]We test for differences using regression analysis. The results are shown in Table 11 in the Appendix.

[28]The problem we address here is that each uninformed player 2 might be facing a player 1 who has chosen the high investment or the low investment with probabilities that differ with the observed outcome. In contrast, a player 2 in treatment FULLINFO can condition her donation on the investment decision of player 1 and she knows that her decision will only become relevant if player 1 has actually chosen the respective investment.



(a) Successful investment. N = 40 in FULLINFO and N = 82 in HIDDENINFO.

(b) Failed investment. N = 48 in FULLINFO and N = 113 in HIDDENINFO.

Figure 2: Mean donation after successful and unsuccessful investments in treatment FULLINFO (imputed with investment distribution from HIDDENINFO) and for uninformed players 2 in treatment HIDDENINFO. Each circle corresponds to one subject-level average. The boxes mark the interquartile range with a bold line at the median. The wide lines indicate the means. N states the number of subjects included.

tions more strongly but the observed differences fall short of statistical significance at conventional levels. This null result goes against the notion that the avoidance of information represents a negative self-selection and does not support a positive selection of more pro-social subjects into information either. Thus, we find no support for Hypothesis 4. However, this result is a weak one; due to sample size, we can rule out only very large effect sizes (see also discussion in Section B). Therefore, we see this part of the analysis as explorative and acknowledge that further studies are needed to better understand possible selection effects.

Result 4. *We do not find statistically significant differences in donations between players 2 in treatment* FULLINFO *and willingly informed players 2 in treatment* HIDDENINFO*.*
Neither do we find statistically significant differences in donations between players 2 in treatment FULLINFO *and willingly uninformed players 2 in treatment* HIDDENINFO*.*

5.2 Investment choices of player 1

We now turn to the behavior of player 1. We expected player 1 to choose the low cost investment more frequently under HIDDENINFO in response to an anticipated increase in selfishness of player 2 when information is hidden (see Hypothesis 5).

In a total of 748 investment decisions, player 1 chose to invest a high amount in 24.1% of the situations. The raw data suggests a treatment difference in the expected direction: The proportion of high cost investments amounts to 30% in treatment FULLINFO, while it is only 21.9% in HIDDENINFO. If we compute the average of all investment decisions for each of the 187 subjects in the role of player 1 and run a two-sided Wilcoxon rank-sum test, we find that these two proportions do not differ significantly (z = -1.786 p = 0.074). The same holds when we consider the investment decisions only from the first round (Pearson χ^2-test, p = 0.109).[29] Thus, we do not find support for Hypothesis 5. This is consistent with the absence of a significant treatment difference in donation behavior (see Result 3).

Result 5. *Player 1 does not choose the high investment significantly more often in treatment* FULLINFO *than* HIDDENINFO.

5.3 Exploratory analysis of subjects' beliefs

After subjects had completed four rounds of investment and donation decisions, we elicited their beliefs about the share of first movers taking the more expensive investment decision in the first round. We only elicit their beliefs about first-round behavior to mitigate the problem that beliefs may be affected by observed behavior over the four rounds.

As illustrated in Figure 3, the treatment neither affects beliefs of subjects in the role of player 1 (two-sided Wilcoxon rank-sum test, z = -1.320, p = 0.188) nor in the role of player 2 (two-sided Wilcoxon rank-sum test, z = -0.225, p = 0.823). In both treatments, subjects in the role of player 1 expect about 40% of the first movers to take the more costly investment decision. For subjects in the role of player 2, the average belief in treatment HIDDENINFO also equals 40%, while the average is slightly higher in the FULLINFO treatment at 46%. Pooling informed players across treatments, we observe that informed players hold relatively precise estimates whereas uninformed players overestimate the share of first movers incurring high investment costs when the investment fails, independent of the outcome of the investment (see Table 2). While only 12.5% of the first movers have invested, the second players believe that almost 40% have invested. Observing a successful investment, second players slightly underestimate the share of investing first movers.

While beliefs are similar across treatments, we find a systematic relationship between beliefs and player 1's own investment decision. In both treatments, first movers who chose a high investment in the first period expect a higher share of investing first movers than those first movers who themselves did not choose the high cost in-

[29] We separately investigate decisions from round 1 because the influence of experience from previous rounds on the investment decisions may vary between the treatments.

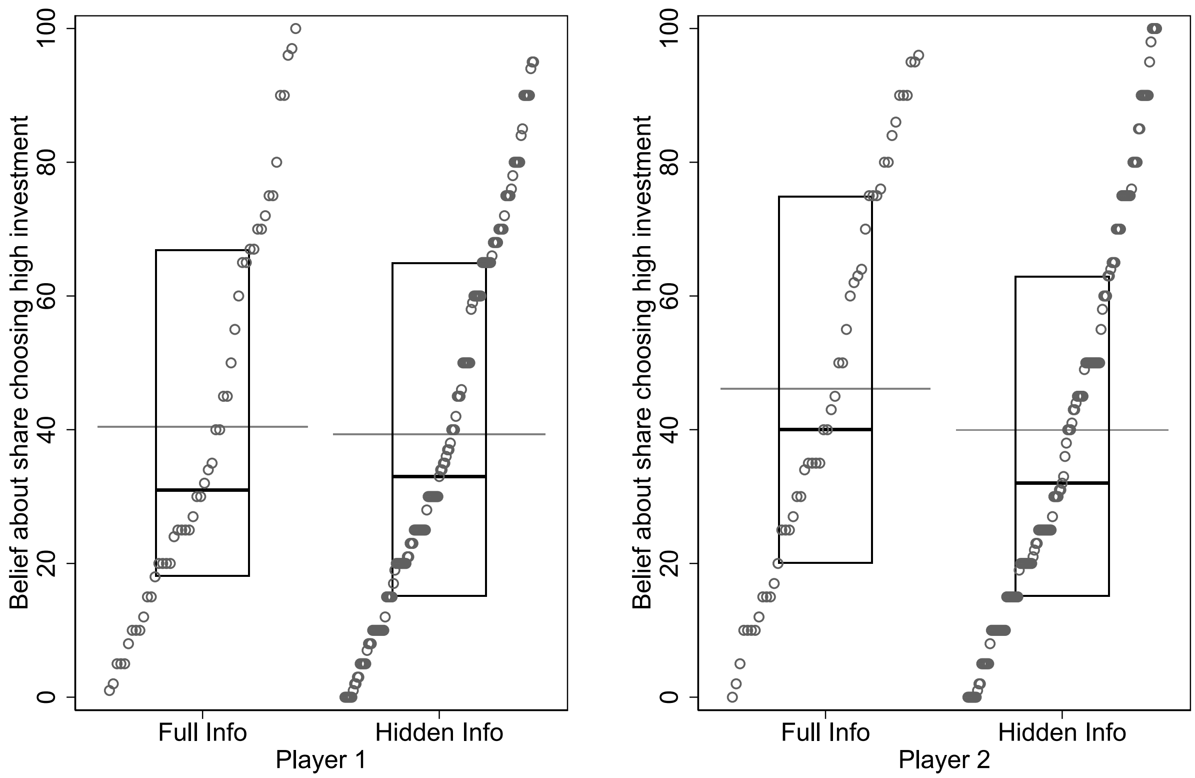


Figure 3: Beliefs about the share of players 1 choosing the high cost investment in round 1, split up by role and treatment. Each circle corresponds to one subject's belief. The boxes mark the interquartile range with a bold line at the median. The wide lines indicate the mean.

vestment. The effect is highly significant and quantitatively large (see Table 15 in the appendix for details).[30]

We further find that subjects in the role of player 2 donate significantly more when holding more optimistic beliefs about the investment decision of player 1. This finding is driven by the beliefs held when the investment was not successful. When the investment succeeded, player 2's donation decision is not systematically related to stated beliefs. For details, see Table 16 in the appendix.

From Table 2, one might conclude that players remain ignorant to maintain their excessively positive beliefs that, in turn, drive their donations. However, this is unlikely to be the case given that the donations of uninformed second movers tend to be lower on average than those of informed players. Instead, second movers might decide to remain ignorant fearing that their positive beliefs are in fact true and would force them to donate more than they are prepared to give. Hence, they seek to avoid certainty about the first mover's decision in order to be able to choose their preferred

[30]Such a false consensus effect is not surprising: subjects expect others to behave like they do, which might also be used as a strategy to justify their own decisions (Ross et al., 1977). See also Engelmann and Strobel (2000) as an early experimental economics study on the false consensus effect and Blanco et al. (2014) on the relevance of false consensus in explaining behavior in social dilemmas.

Table 2: Beliefs and investments in treatment HIDDENINFO

	All	Informed	Uninformed
All decisions			
Investments	0.277	0.390	0.229
Beliefs	0.399	0.409	0.395
All decisions with a successful investment			
Investments	0.500	0.600	0.438
Beliefs	0.461	0.561	0.399
All decisions with an unsuccessful investment			
Investments	0.141	0.190	0.125
Beliefs	0.362	0.265	0.393

Notes: For this table, we consider those 41 subjects as informed who acquired information about player 1's investment in round 1. Out of these 20 faced a successful investment (high outcome) and 21 a failed investment (low outcome) when deciding about their information choice.

donation. Moreover, holding positive beliefs may provide utility to individuals so that they might be optimally off.

6 Concluding discussion

In principal-agent situations, reciprocity may play an important role, i.e. the principal may react to a pro-social act of the agent with more generosity than to a selfish action. However, the agent may only have limited control over the consequences of his action since other influencing factors may also play a role. When deciding how to react toward the agent, the principal can decide how much weight to give to the chosen actions and the realized outcome. Thus far, evidence in the literature supports the importance of intentions (Charness and Levine, 2007). However, it relies on the fact that the principal observes both outcome and intention – an assumption that may be unrealistic in many situations.

In this paper, we contribute to this literature with results from a design where outcomes are always available whereas intentions may be hidden. We use a controlled laboratory experiment on a two-player sequential interaction to investigate whether behavior of the principal (player 2) — a dictator game donation — depends on whether the agent's (player 1's) intention — an investment choice that benefits both players — is observed or may only be revealed by player 2 upon paying a symbolic fee. We also ask if the behavior of player 1 changes with the visibility of his action.

Our paper extends the literature on the exploitation of moral wiggle room in social decisions to principal-agent-settings. While similar to other studies on the exploitation of moral wiggle room in the context of reciprocal actions (see, e.g., Regner, 2018;

Van der Weele et al., 2014; Regner and Matthey, 2021), our focus lies on the reciprocation behavior of the principal. In the typical experiments studying the exploitation of moral wiggle room through information avoidance, the decision maker can resolve uncertainty about the consequences of her actions on the other party. In our setting, in contrast, the decision maker can resolve uncertainty about the *deservingness* of the agent. In other words, the principal can gather information whether the agent deserves a generous donation, while, in contrast to the typical studies, she is always aware of the consequences of her actions for the agent. Our study thus investigates another dimension of missing information in two-player-interactions.

Our results show that the first mover's intention has a large positive impact on the second mover's generosity, while the outcome has no effect. Despite donations responding strongly to the investment decision if known, hiding the investment decision by default does not appear to have a detrimental effect on donations. While we find no significant evidence that subjects' choice of information correlates with their inclination to be more generous toward player 1 after a good outcome, the point estimates have the expected signs and we only have limited sensitivity. Therefore, we suggest that our null result should be interpreted as ruling out only very large effect sizes in this part of the analysis.

The result that donations under full and under hidden information do not differ significantly appears surprising at first, given that moral wiggling is often observed when outcomes are hidden. We offer two explanations why this might be different in the case of intentions. First, we believe that the first movers' intentions in our study provide less moral wiggle room to the second mover than outcomes in previous experiments on moral wiggle room. Recall the typical binary decision situations used to study moral wiggle room where decision makers who would have chosen the prosocial option under full information but choose the selfish option under hidden information are said to exploit moral wiggle room. These individuals exhibit pro-social preferences of an intermediate intensity. Those with stronger pro-social preferences choose the altruistic option under both information conditions, whereas subjects with weaker pro-social preferences always choose the egoistic option. If a substantial share of subjects exhibit intermediate pro-social preference intensities, exploitation of moral wiggle room can be detected. Our result could be a consequence from fewer subjects having these intermediate preference intensities with respect to the two-player interaction used here. One important difference between wiggling with respect to outcome or intention is that the first mover, the one affected by the second player's decision is not a third party but is related to the second player through his own choice, his intention. Given that we do not find evidence for the exploitation of moral wiggle room, we surmise that subjects' pro-social preferences with respect to reciprocal interactions are different from those that are at play when it comes to the side-effects of their decisions on a third party.

An alternative explanation for donations being very similar across treatments lies in the possibly important role of beliefs. Players in HIDDENINFO seem to not hold rational beliefs but they tend to be too pessimistic about the first mover's investment after successes and to be too optimistic after failures, on average (see Section 5.3). Thus, our data is also consistent with the idea that part of our uninformed subjects do wiggle but given that their beliefs are too optimistic after failures and too pessimistic after successes, this does not lead to them giving less than what they would have given on average if informed.

By shedding light on potential wiggling with respect to others' intention, our study provides a new perspective to the literature. Previous studies on moral wiggle room in the context of reciprocity have found opposite results: The findings in Regner (2018) support the notion that moral wiggle room carries over from dictator games to the richer context of trust games, whereas the results in Van der Weele et al. (2014) contradict this notion. A reason for these opposing findings may be a ceiling effect as decisions in the baseline in Van der Weele et al. (2014) are already very selfish. Our paper adds to this strand of literature by lending support to the findings in Van der Weele et al. (2014) and showing that in settings that are less abstract than the dictator game paradigm with an affected third party, wiggling may be less prevalent. Further, we note that our experimental design differs significantly from the ones in Regner (2018) and Van der Weele et al. (2014) and as such contributes an additional perspective to the literature on moral wiggling. We create the possibility to exploit moral wiggle room through hidden information, whereas they implement the other treatment variations of the seminal paper by Dana et al. (2007): plausible deniability (Van der Weele et al., 2014; Regner, 2018) and multiple dictators (Regner, 2018).

Based on our experiment and its results, we infer that cooperation does not need to suffer when information on the agent's intention is not readily observable to the principal. While intentions can indeed only be rewarded when they are known to the principal, we observe that uninformed principals hold unrealistically positive beliefs about the agent's intentions and do not seem to take into account the information from the observed outcome. This goes against any possible negative effect from being uninformed about the agent's intentions. Consistent with Ging-Jehli et al. (2020), principals do not seem to be strategically cynical about the agent's behavior but rather exhibit an optimism bias. This also implies that ignorance is apparently not used to blame the agent as could be thought following Gurdal et al. (2013). Additional research is needed to better understand how beliefs of both players are formed, how they relate to subject's preferences and information as well as whether beliefs influence behavior or *vice versa*.

References

Alesina, A. and G.-M. Angeletos (2005). Fairness and redistribution. *American Economic Review 95*(4), 960–980.

Andreoni, J. and J. H. Miller (1993). Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *The Economic Journal 103*(418), 570–585.

Bartling, B., E. Fehr, and K. M. Schmidt (2013). Jeea-fbbva lecture 2012: Use and abuse of authority: A behavioral foundation of the employment relation. *Journal of the European Economic Association 11*(4), 711–742.

Bénabou, R. and J. Tirole (2011). Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics 126*(2), 805–855.

Blanco, M., D. Engelmann, A. K. Koch, and H.-T. Normann (2014). Preferences and beliefs in a sequential social dilemma: A within-subjects analysis. *Games and Economic Behavior 87*, 122–135.

Brandts, J. and G. Charness (2011). The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics 14*(3), 375–398.

Brownback, A. and M. A. Kuhn (2019). Understanding outcome bias. *Games and Economic Behavior 117*, 342–360.

Charness, G. and D. I. Levine (2007). Intention and stochastic outcomes: An experimental study. *The Economic Journal 117*(522), 1051–1072.

Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics 117*(3), 817–869.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics 14*(1), 47–83.

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

Engelmann, D. and M. Strobel (2000). The false consensus effect disappears if representative information and monetary incentives are given. *Experimental Economics 3*(3), 241–260.

Engelmann, D. and M. Strobel (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review 94*(4), 857–869.

Erkal, N., L. Gangadharan, and B. H. Koh (2021). By chance or by choice? biased attribution of others' outcomes when social preferences matter. *Experimental Economics*, 1–31.

Exley, C. L. (2016). Excusing selfishness in charitable giving: The role of risk. *The Review of Economic Studies 83*(2), 587–628.

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

Faul, F., E. Erdfelder, A.-G. Lang, and A. Buchner (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods 39*(2), 175–191.

Fehr, E. and K. M. Schmidt (2006). The economics of fairness, reciprocity and altruism–experimental evidence and new theories. *Handbook of the Economics of Giving, Altruism and Reciprocity 1*, 615–691.

Feiler, L. (2014). Testing models of information avoidance with binary choice dictator games. *Journal of Economic Psychology 45*, 253–267.

Festinger, L. (1957). *A theory of cognitive dissonance*, Volume 2. Stanford university press.

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

Ging-Jehli, N. R., F. H. Schneider, and R. A. Weber (2020). On self-serving strategic beliefs. *Games and Economic Behavior 122*, 341–353.

Golman, R., D. Hagmann, and G. Loewenstein (2017). Information avoidance. *Journal of Economic Literature 55*(1), 96–135.

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

Grossman, Z. (2014). Strategic ignorance and the robustness of social preferences. *Management Science 60*(11), 2659–2665.

Grossman, Z. and J. J. Van Der Weele (2017). Self-image and willful ignorance in social decisions. *Journal of the European Economic Association 15*(1), 173–217.

Gueth, W., S. Huck, and P. Ockenfels (1996). Two-level ultimatum bargaining with incomplete information: An experimental study. *The Economic Journal 106*(436), 593–604.

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

Güth, W. and M. G. Kocher (2014). More than thirty years of ultimatum bargaining experiments: Motives, variations, and a survey of the recent literature. *Journal of Economic Behavior & Organization 108*, 396–409.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, and R. McElreath (2001). In search of homo economicus: Behavioral experiments in 15 small-scale societies. *American Economic Review 91*(2), 73–78.

Kajackaite, A. (2015). If I close my eyes, nobody will get hurt: The effect of ignorance on performance in a real-effort experiment. *Journal of Economic Behavior & Organization 116*, 518–524.

Kandul, S. (2016). Ex-post blindness as excuse? The effect of information disclosure on giving. *Journal of Economic Psychology 52*, 91–101.

Larson, T. and C. M. Capra (2009). Exploiting moral wiggle room: Illusory preference for fairness? A comment. *Judgment and Decision Making 4*(6), 467.

Ledyard, J. O. (1994). Public goods: A survey of experimental research. California Institute of Technology.

Lind, J. T., K. Nyborg, and A. Pauls (2019). Save the planet or close your eyes? Testing strategic ignorance in a charity context. *Ecological Economics 161*, 9–19.

McCabe, K. A., M. L. Rigdon, and V. L. Smith (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization 52*(2), 267–275.

Momsen, K. and M. Ohndorf (2020). When do people exploit moral wiggle room? an experimental analysis of information avoidance in a market setup. *Ecological Economics 169*, 106479.

Momsen, K. and M. Ohndorf (2021). Information avoidance, selective exposure, and fake (?) news: Theory and experimental evidence on green consumption. *Journal of Economic Psychology*, 102457.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica 46*(1), 69–85.

Pan, X. and E. Xiao (2016). It's not just the thought that counts: An experimental study on the hidden cost of giving. *Journal of Public Economics 138*, 22–31.

Regner, T. (2018). Reciprocity under moral wiggle room: Is it a preference or a constraint? *Experimental Economics 21*(4), 779–792.

Regner, T. and A. Matthey (2021). Actions and the self: I give, therefore i am? *Frontiers in psychology 12*.

Ross, L., D. Greene, and P. House (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology 13*(3), 279–301.

Schaffer, M. E. and S. Stillman (2006). Xtoverid: Stata module to calculate tests of overidentifying restrictions after xtreg, xtivreg, xtivreg2, xthtaylor.

Shang, J. and R. Croson (2009). A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *The Economic Journal 119*(540), 1422–1439.

Toussaert, S. (2017). Intention-based reciprocity and signaling of intentions. *Journal of Economic Behavior & Organization 137*, 132–144.

Van der Weele, J. J., J. Kulisa, M. Kosfeld, and G. Friebel (2014). Resisting moral wiggle room: how robust is reciprocal behavior? *American Economic Journal: Microeconomics 6*(3), 256–64.

Appendix

A Descriptive statistics

Table 3: Descriptive Statistics

	Hidden			Full			
	#Obs.	Mean	SD	#Obs.	Mean	SD	p-value
Female	274	0.47	0.50	100	0.42	0.50	0.418
Male	274	0.52	0.50	100	0.55	0.50	0.586
Age	274	23.0	4.21	100	23.6	5.76	0.921
Trust	274	1.44	0.63	100	1.55	0.58	0.134
Management & Economics	274	0.18	0.38	100	0.25	0.44	0.126
Social Sciences	274	0.084	0.28	100	0.12	0.33	0.289
Engineering Sciences	274	0.28	0.45	100	0.28	0.45	0.960
Education Sciences	274	0.018	0.13	100	0.030	0.17	0.487
Law	274	0.073	0.26	100	0.040	0.20	0.249
Psychology	274	0.0036	0.060	100	0.010	0.10	0.456
Natural Sciences	274	0.16	0.36	100	0.050	0.22	0.006
Sports	274	0.051	0.22	100	0.070	0.26	0.482
Medicine	274	0.0073	0.085	100	0.020	0.14	0.291
Other	274	0.15	0.36	100	0.13	0.34	0.633

Notes: Sample characteristics split by treatment. Subjects self-classified as Male, Female, Diverse, or other so that shares of male and female subjects do not add up to 100%. P-values refer to a two-sided Wilcoxon rank-sum test for "Age" and "Trust" and to a χ^2-test for all the other variables.

Table 4: Number of decisions for given investment and outcome, number of subjects in the role of player 2 who took at least one of the described decisions, and average number of those decisions per subject. Investment in HIDDENINFO by default unknown to subjects.

	investment failed			investment succeeded		
	decisions	subjects	dec./subj.	decisions	subjects	dec./subj.
FULLINFO (strategy method)						
low investment	118	48	2.46	82	40	2.05
high investment	118	48	2.46	82	40	2.05
HIDDENINFO (direct response method)						
either investment	352	135	2.61	196	107	1.83
low investment	322	133	2.4	106	79	1.34
high investment	30	29	1.03	90	70	1.29

To better understand the numbers in the table, let us focus on the subjects in the HIDDENINFO treatment facing a failed investment. Out of the 187 subjects in the role of player 2 in treatment HIDDENINFO, 135 or 98.5% took at least one decision where the investment of player 1 had failed. Of these subjects, each took on average 2.61 decisions of this type. 133 subjects (97.1%) were at least once in a situation where the low cost investment had failed and 29 (21.2%) faced at least once a situation where the high cost investment had failed. Clearly, these numbers cannot add up to the count of subjects facing at least once a failed investment, as some subjects face both an unsuccessful low cost and high cost investment, whereas others are confronted with only one kind of investment preceding a failure. Note, however, that player 2 may or may not have known about the investment decision of player 1 depending on her decision to learn or to avoid this information, an issue that we discuss later. The last two rows in Table 4 are therefore only of theoretical interest as they contain both informed and uninformed players.

Table 5: Average donation of player 2 in treatment FULLINFO. Standard deviations in parentheses. Unit of observation is the subject-level average.

	investment failed	investment succeeded
low cost investment	1.93 (3.58)	2.49 (4.36)
high cost investment	5.46 (5.41)	6.85 (5.39)
No. of subjects	48	40

Table 6: Conditional empirical frequency of high investment in HIDDENINFO.

	Cond. frequ. of high investment	Percent of decisions
High outcome	0.46	64%
Low outcome	0.09	36%

Table 7: Average donations in HIDDENINFO and FULLINFO.

	Hidden Info	Full Info (imputed)
High outcome	4.79 (6.09)	4.49 (4.22)
No. of subjects	107	40
Low outcome	2.76 (4.45)	2.23 (3.52)
No. of subjects	135	48

Notes: We use the empirical distribution of high and low investments conditional on each outcome in HIDDENINFO from Table 6 to impute hypothetical means in FULLINFO. Observations are subject level averages. Standard deviations in parentheses.

B Power and sensitivity analysis

As documented in the preregistration, we had planned to collect data from a total of 20 sessions, thereof five in treatment FULLINFO and 15 in HIDDENINFO. The session were planned to be split across two labs. In the lab at TU Berlin, we had planned to run 12 sessions with 22 participants per session. Of these sessions, three and thus 66 participants were planned to be in FULLINFO and nine session with 198 participants in HIDDENINFO. For the PLEx in Potsdam we had planned eight sessions with 20 participants per session. Two of these session were planned to be run as FULLINFO with 40 participants in total, and another six sessions with a total of 120 participants were planned to be run in HIDDENINFO. To summarize, we had planned with 106 participants in FULLINFO and 318 participants in HIDDENINFO, which would have given us 53 participants in each role in FULLINFO and 159 participants in each role in HIDDENINFO.

Our realized sample comprises 100 participants in FULLINFO and 274 participants in HIDDENINFO, resulting in 50, respectively 137 participants in each role.

For Hypotheses 1 and 2, we use two-sided Wilcoxon signed rank tests to asses the within-subject differences in donations due to intentions and outcomes in the treatment FULLINFO. Using G*Power (Faul et al., 2007), we compute that with the planned sample size of 53 subjects in the role of player in FULLINFO, we would be able to detect effect sizes of at least 0.4 at a power of 0.8 and with $\alpha = 0.05$. The realized sample of 50 is smaller than planned. We show below that our analysis is still powered to detect medium to large effects in Results 1 and 2 even though it is less sensitive than originally planned.

Hypothesis 1 specifies a comparison conditional on the investment outcome. Not every player 2 took a decision for both a low and a high outcome. Taking this into account, we end up with a sample of 48 for comparisons for the test conditioning on a low outcome and 40 for comparisons for the test conditioning on a high outcome. A sensitivity analysis reveals that we would have picked up an effect with Cohen's $d = 0.42$ at a power of 0.8 and $\alpha = 0.05$ for the first comparison testing whether—conditional on a failed investment—donations are higher for high than for low investments. For the second comparison testing whether—conditional on a successful investment—donations are higher for high than for low investments, a sensitivity analysis shows that effects with Cohen's $d = 0.47$ can be detected with a power of 0.8 and $\alpha = 0.05$.

Hypothesis 2 specifies a comparison conditional on the investment decision. As we use the strategy method in FULLINFO, player 2 takes decisions for both the high and the low investment case. However, we can only do the comparison between high and low outcome conditional on an investment choice for players who at least once faced a high and a low outcome. This reduces our sample to 38 subjects. At a power of 0.8

with $\alpha = 0.05$, the Wilcoxon signed rank test behind Result 2 would pick up an effect with Cohen's $d = 0.48$.

For Hypotheses 3 and 4, we use two-sided Wilcoxon-Mann-Whitney tests that allow us to compare donations of player 2 across the treatments FULLINFO and HIDDENINFO. Already with the planned sample size we would have been able to detect only large effect sizes at a power of 0.8 and the actual sample is even smaller than planned so that the analysis is even less sensitive, see details below. But the observed effect size is so low that lack of power alone is unlikely to drive the finding that donations are not significantly lower on average in HIDDENINFO than in FULLINFO.

To test Hypothesis 3, we compare average donations conditional on either investment outcome across treatments. Conditional on the high outcome, we can use data from 40 subjects in FULLINFO and from 107 subjects in HIDDENINFO. At a power of 0.8 and with $\alpha = 0.05$, we would detect effects with a Cohen's d of 0.54 or larger. But the observed effect size amounts only to a Cohen's d of 0.053. Conditional on the low outcome, we can use data from 48 subjects in FULLINFO and from 135 subjects in HIDDENINFO. A sensitivity analysis shows that, at a power of 0.8 and with $\alpha = 0.05$, we would detect effects with a Cohen's d of 0.48 or larger. Again, the observed effect size is much smaller with a Cohen's d of 0.132. With the originally planned sample of 53 subjects in FULLINFO and 159 in HIDDENINFO, assuming that we had observed decisions for low and high outcomes for each of these participants, we would have been able to pick up effects with $d = 0.46$ or larger at a power of 0.8 and $\alpha = 0.05$. Thus, the observed effect sizes would have been found insignificant also at the originally planned sample size and to detect such small effect sizes, a much larger sample would be needed.

The picture is a different one in the case of Hypothesis 4. To be well-powered to test whether informed players 2 in HIDDENINFO choose donations that are different from those of players 2 in FULLINFO, we would need a larger sample. A major problem here is that player 2 self-selects into information and only a quarter of decisions are endogenously informed. Under the best circumstances (which have not realized) this would imply that for every participant in the role of player 2, we have one informed decision. Assuming again the best circumstances, these informed decisions would divide equally into those where player 1 chose a high and a low investment, respectively. This would imply that in HIDDENINFO the best constellation would have comprised 78 subjects who faced a high outcome, thereof 59 would have been matched with a player 1 who had chosen a high investment and 19 with a player 1 who had chosen a low investment. Another 78 subjects would have faced a low outcome and there 59 would have been matched with a player 1 who had chosen a low investment and 19 with a player 1 who had chosen a high investment. Under these most favorable conditions for the planned sample size, our analysis using two-sided Wilcoxon-Mann-Whitney tests – targeting $\alpha = 0.05$ and a power of 0.8 – would have a sensitivity of $d = 0.56$ for the comparison of informed players across treatments conditional on a successful high investment and for the same comparison conditional on a failed low investment, and a sensitivity of $d = 0.79$ conditional on a successful low investment and conditional on a failed high investment. Thus, we admit, that we were ex ante only well-powered to compare donations for the more likely constellations of the successful high investment and the failed low investment.

Ex post, we see that we did not only collect fewer observations than planned but these also do not distribute well over the conditions and into information conditions. Thus, for the realized data, with $\alpha = 0.05$ and a power of 0.8, we have only a sensitivity of $d = 0.74$ for the comparison of informed players across treatments conditional on a successful high investment, of $d = 0.76$ for the same comparison conditional on a successful low investment, a sensitivity of $d = 1.12$ conditional on a failed high investment, and a sensitivity of $d = 0.60$ for the comparison of informed donations across treatments conditional on a failed high investment. Thus, we do only have the power to detect large to very large effects. Thus, the insignificance of the non-parametric tests tells us that possible difference are not large but we cannot, using these tests, rule out with confidence that there are medium sized to large or small effects. We believe that the insignificance of our tests on Hypothesis 4 should therefore be taken cautiously and additional studies are needed to better understand self selection into information based on social preferences.

For Hypothesis 5, the two-sided Wilcoxon-Mann-Whitney tests to test for treatment differences in investments of player 1 with 50 subjects in FULLINFO and 137 in HIDDENINFO has, imposing a power of 0.8 and $\alpha = 0.05$, a sensitivity of $d = 0.48$. The observed effect size is much smaller with a standardized effect size of only 0.263. Ex post power is therefore relatively low with only 0.34 and we cannot rule out that there is a small difference in investment behavior that we are not powered to detect with our study. This is not a result of the study being smaller than planned but the original sample size would have yielded a sensitivity of $d = 0.46$ still way above the realized effect size.

C Additional tables, figures and analyses

Table 8: Regression outcome: Player 2 donations regressed on player 1 investment outcome and investment decision (basis for Table 1)

	Full Information	Willingly Informed Hidden Information
Investment	3.80***	9.24**
	(0.751)	(3.442)
Success	0.69	1.46
	(0.564)	(0.944)
Investment∗Success	-0.09	-3.53
	(0.794)	(3.995)
Period	-0.24*	0.46
	(0.118)	(0.295)
Constant	2.30***	1.14
	(0.534)	(1.135)
Observations	400	148
No. of subjects	50	63

Notes: Dependent variable is player 2 donation. *Investment* and *Success* are dummies for the investment having been high and successful respectively. Output from random-effects regressions. Column 1 uses data from FULLINFO, column uses data from informed players in HIDDENINFO. Cluster-robust standard errors in parentheses (clustered on subject-level).

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

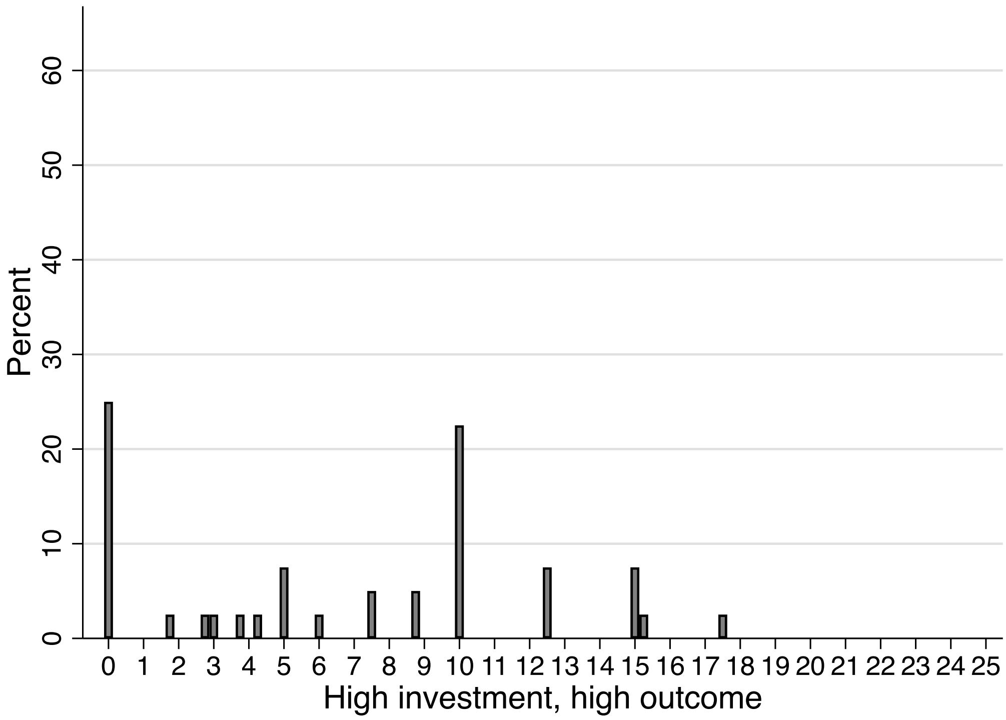


Figure 4: Histogram of subject-level average donations conditional on a high investment and a high outcome (FULLINFO)

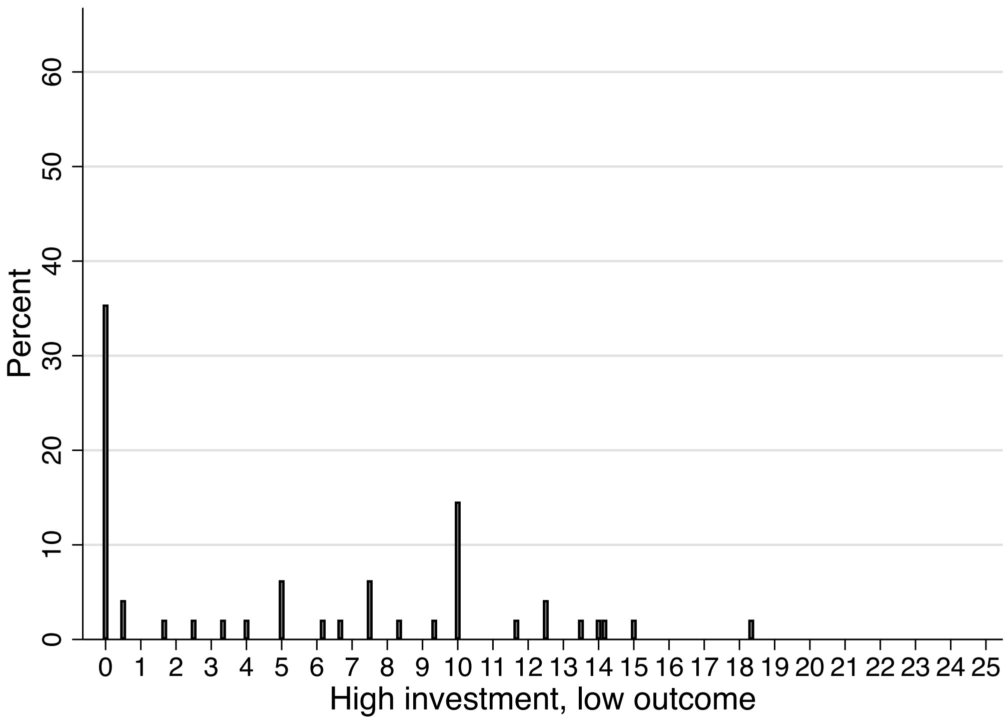


Figure 5: Histogram of subject-level average donations conditional on a high investment and a low outcome (FULLINFO)

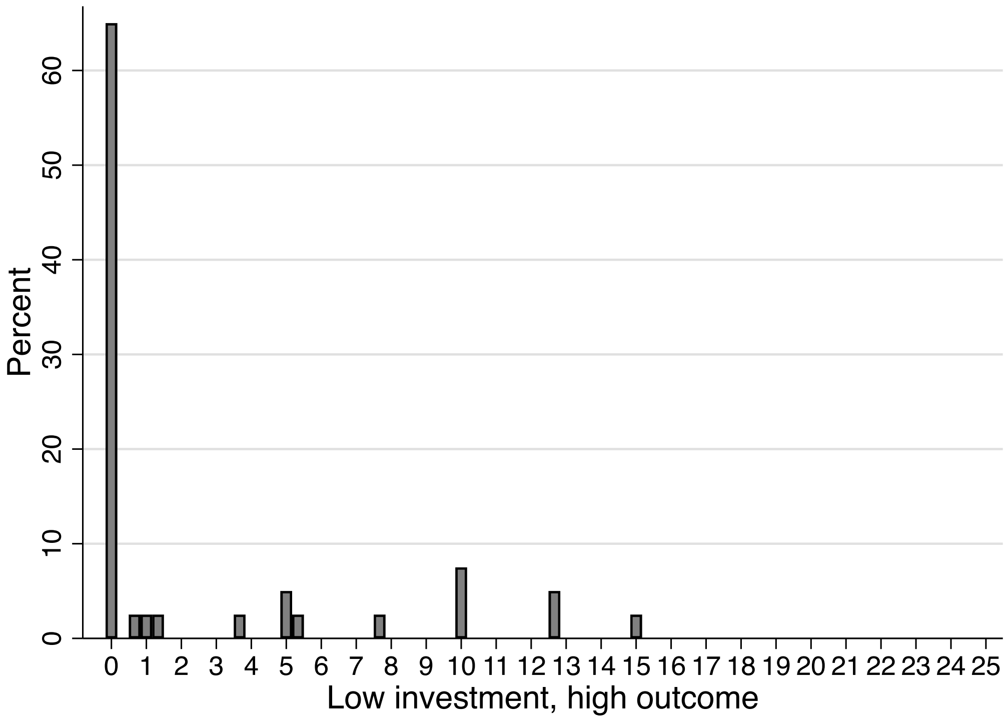


Figure 6: Histogram of subject-level average donations conditional on a low investment and a high outcome (FULLINFO)

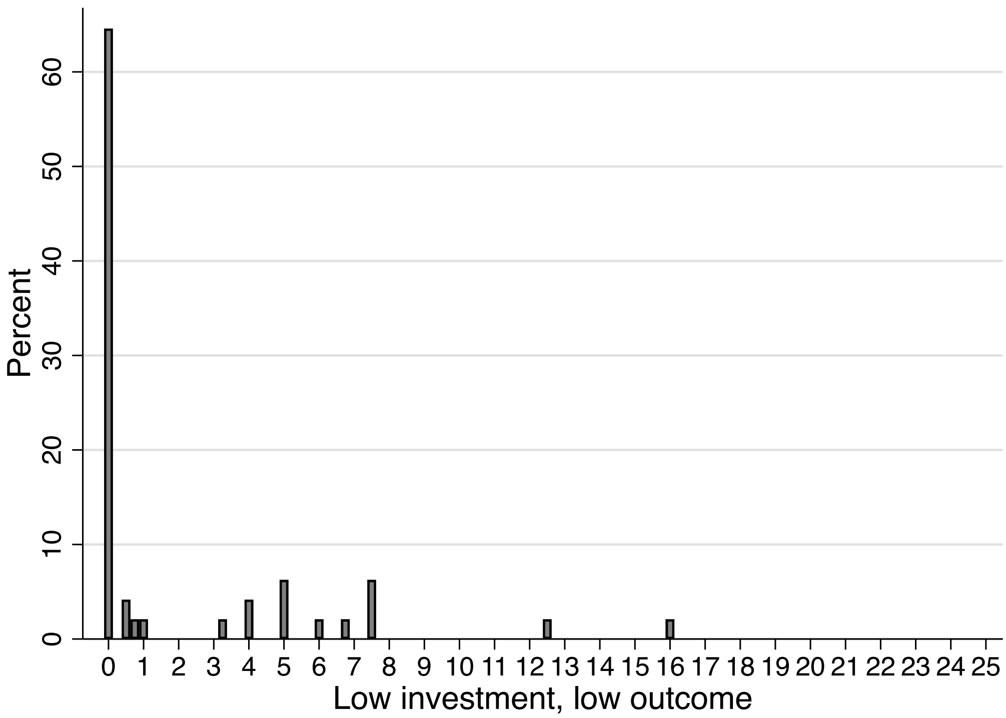


Figure 7: Histogram of subject-level average donations conditional on a low investment and a low outcome (FULLINFO)

Table 9: Regression outcome: Player 2 donations regressed on treatment condition.

	All data pooled
Full	-1.00
	(0.629)
Investment	1.67
	(1.220)
Success	1.67***
	(0.465)
Full*Success	-0.87
	(0.721)
Full*Investment	2.13
	(1.430)
Success*Investment	-0.53
	(1.367)
Full*Inv.*Suc.	0.44
	(1.579)
Period	-0.10
	(0.100)
Constant	2.90***
	(0.426)
Observations	948
No. of subjects	187

Notes: Dependent variable is player 2 donation. *Investment* and *Success* are dummies for the investment having been high and successful respectively. *Full* is a dummy for the treatment with full information. Output from random-effects regressions. Cluster-robust standard errors in parentheses (clustered on subject-level). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 10: Marginal effect from regressions of player 2 donations regressed on treatment condition (see Table 9 in the main text)

	(1)
Full	
Success =0	-0.28
	(0.691)
Success =1	-1.00
	(0.799)
Investment =0	-1.33*
	(0.639)
Investment =1	0.96
	(1.071)

Notes: Dependent variable is player 2 donation. *Investment* and *Success* are dummies for the investment having been high and successful respectively. *Full* is a dummy for the treatment with full information. Marginal effects from random-effects regressions as summarized in Table 9. Cluster-robust standard errors in parentheses (clustered on subject-level). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 11: Regression of player 2 information choice on investment outcome.

	InfoIntention
Success	0.03
	(0.036)
Period	-0.02
	(0.013)
Constant	0.32***
	(0.048)
Observations	548
No. of Subjects	137

Notes: Dependent variable is player 2′ decision to acquire information (1) or not (0). *Success* is a dummy for the investment of player 1 having been successful. Results from random effects regressions. Cluster-robust standard errors in parentheses (clustered on subject-level). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 12: Comparing player 2 donations across treatments, separately for informed and uninformed players in treatment HIDDENINFO

	Willingly Informed	Uninformed
Full	-0.72	0.89
	(0.960)	(0.686)
Investment	8.50*	
	(3.313)	
Success	1.28	1.46**
	(0.952)	(0.534)
Full$*$Success	-0.45	-0.74
	(1.094)	(0.661)
Full$*$Investment	-4.70	
	(3.398)	
Investment$*$Success	-2.89	
	(3.846)	
Full$*$Inv.$*$Suc.	2.81	
	(3.928)	
Period	-0.07	-0.14
	(0.117)	(0.097)
Constant	2.55**	3.03***
	(0.886)	(0.474)
Observations	548	800
No. of subjects	113	172

Notes: Dependent variable is player 2 donation. In column 1, we compare decisions in FULLINFO with decisions by willingly informed players in HIDDENINFO. In column 2, we compare decisions in FULLINFO with decisions by uninformed players in HIDDENINFO. *Full* is a dummy taking the value 1 for observations in treatment FULLINFO and 0 for those in HIDDENINFO. *Success* is a dummy for the investment having been successful. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 13: Player 2 donations conditional on player 1 investment outcome and investment decision respectively controlling for investment's outcome in t-1

	Full Info		Willingly Informed Hidden Info	
	(1)	(2)	(3)	(4)
Investment				
Success=0	4.052***		11.917*	
	(0.815)		(5.029)	
Success=1	2.736***		5.940**	
	(0.759)		(1.563)	
Success				
Investment=0		1.3158		1.445
		(0.7758)		(1.334)
Investment=1		0.0001		-4.532
		(0.7914)		(5.540)
Observations	300	300	107	107
No. of Subjects	50	50	54	54

Notes: Dependent variable is player 2 donation. *Investment* and *Success* are dummies for the investment having been high and successful respectively. Output from random-effects regressions (marginal effects). Columns 1 and 2 report the effect of *Investment* conditional on *Success* being 0 or 1 and viceversa respectively in the Full information treatment. Column 2 and 3 report the effect of *Investment* conditional on *Success* being 0 or 1 and vice versa respectively in the Hidden information treatment among willingly informed players 2. Cluster-robust standard errors in parentheses (on subject-level).
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 14: Player 2 donations conditional on player 1 investment decision split by inequality aversion attitudes only for the Berlin sample

	(1) Overall	(2) H outcome	(3) L outcome	(4) Overall	(5) H outcome	(6) L outcome
Investment						
Inequality=0	4.239***	4.054***	4.364***			
	(0.894)	(1.011)	(1.097)			
Inequality=1	2.719*	3.615*	2.105			
	(1.332)	(1.411)	(1.534)			
Unemployment=0				4.219***	4.306***	4.167***
				(0.886)	(1.094)	(1.045)
Unemployment=1				2.571*	3.000***	2.143
				(1.299)	(0.900)	(1.946)
Observations	248	100	148	248	100	148

Notes: Dependent variable is player 2 donation. *Investment* is a dummy for the investment having been high or low cost. Output from random-effects regressions (marginal effects). *Inequality* and *Unemployment* are dummies that assume value 0 for people who answered 3, 4 or 5 on a scale from 1 to 5 where 5 means "I fully agree" to the state fighting income inequality and unemployment, respectively. Column 1 reports the effect of *Investment* conditional on *Inequality* being 0 or 1. Column 2 reports the effect of *Investment* conditional on *Unemployment* being 0 or 1. Cluster-robust standard errors in parentheses (on subject-level).

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 15: Beliefs of first movers regarding investment decisions

	(1) No controls	(2) Controls
Investment	30.120***	30.576***
	(3.641)	(3.815)
Full	-2.560	-2.849
	(3.870)	(4.025)
Constant	30.952***	25.802***
	(2.545)	(9.776)
Controls	No	Yes
Observations	187	187

Notes: OLS. Dependent variable is the belief regarding the share of first movers choosing the costly investment. *Full* is a dummy for treatment FULLINFO and *Investment* is a dummy for own investment in round 1 being high. Controls are age, a dummy for male and a dummy for studying business or economics. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 16: Donations of uninformed player 2 in treatment HIDDENINFO

	(1) all	(2) successful	(3) unsuccessful
Belief	0.041*	0.044	0.037*
	(0.017)	(0.048)	(0.017)
Constant	-4.440	-3.651	-3.107
	(3.386)	(5.419)	(3.655)
Controls	Yes	Yes	Yes
Observations	96	32	64

Notes: Dependent variable is the donation of an uninformed player 2 in period 1. Output from an OLS regression model. Controls are the subjects' age, a dummy for being male and a dummy for studying business or economics. .* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

D Instructions (translated from German)

Welcome to our experiment!

The experiment you will now participate in is designed to analyze economic decision making behavior. In this experiment you can earn money and the amount you will receive in the end depends on the decisions you and other participants make. The amount of your payout at the end of the experiment also depends on how well you have understood the following instructions. All statements in the instructions are true and the instructions are identical for all participants. Please read the instructions carefully now.

During the experiment you are not allowed to use electronic devices or communicate with other participants. Please use only the programs and functions intended for the experiment. Please do not talk to the other participants. If you have a question, please raise your hand. We will then come to you and answer your question in silence. Please do not ask your questions out loud under any circumstances. If the question is relevant for all participants, we will repeat it aloud. If you violate these rules, we will have to exclude you from the experiment and payout.

Please read these instructions carefully now. The instructions are identical for all participants.

Today's experiment consists of two parts. These instructions refer to the first part of the experiment. Instructions for the second part will be displayed on your screen once the first part is complete. The two parts are completely independent and your earnings from the experiment are calculated from your earnings in the two parts.

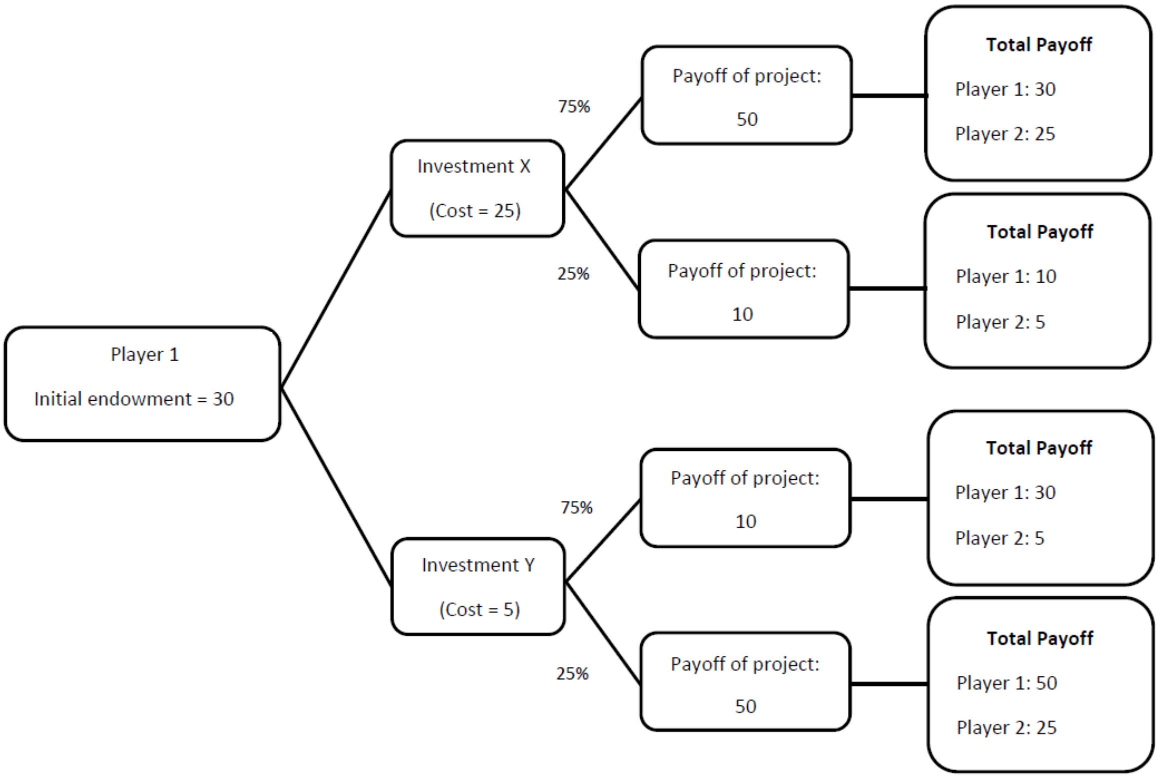
The first part of the experiment consists of 4 independent rounds. Only one of these rounds is relevant for payment. Which one it will be is determined randomly at the end of the experiment.

There are two types of players in the experiment, player 1 and player 2. Which role you play is determined randomly. You keep your role for all rounds of the experiment. You will be divided into groups of two, each consisting of player 1 and player 2. In each round, new groups of two are randomly formed, so you will only interact with the same player once. You will never know the identity of your fellow players. In each round, Player 1 decides first (Decision Phase 1), followed by Player 2 (Decision Phase 2).

Decision stage 1

Player 1 has an initial endowment of 30 points. From these 30 points he can invest either 5 or 25 points in a common project. This project affects the payouts of both players in a group of two, as the amount in the project is divided equally between both players. Player 1 can choose between two investment options. Both investment options can either succeed (50 points) or fail (10 points). They differ in the probability of success or failure. They also have different costs for player 1.

The investment opportunities for player 1 are as follows:



- **Investment X:** This investment costs player 1 25 points. With a probability of 75% it will be successful, i.e. it will lead to the high payout of 50 points from the project, with a probability of 25% it will fail (10 points).
- **Investment Y:** This investment costs player 1 5 points. With a probability of 25% it will be successful, i.e. it will lead to a high payout of 50 points from the project, with a probability of 75% it will fail (10 points).

Player 1 and Player 2 each receive half of the final amount in the project. If the investment was successful, each player gets 25 points, if it was not successful, each player gets 5 points from the project. Note that player 1 has kept either 5 points (Investment Y) or 25 points (Investment X) from his initial investment.

Decision stage 2

After Player 1 has made his investment decision for the joint project, it is Player 2's turn. Player 2 receives an amount of 30 points. This amount is independent of the payout from decision phase 1. 25 of these 30 points can be split between player 1 and himself. Any split is possible; he can, for example, keep the entire 25 points for himself, share them with player 1 (e.g. 12.50 points for each player) or transfer the initial endowment completely to player 1.[31]

only in FULLINFO*:*
[Player 2 observes the result of the investment decision without knowing whether player 1 has chosen Investment X or Investment Y. Player 2 now makes two decisions: In case player 1 has chosen Investment X, he has to decide how he would divide the 25 points between himself and player 1. He also has to decide how he would split his 25 points if player 1 chose Investment Y. *After he has made both decisions, the actual decision made by player 1 determines which of the decisions of player 2 is implemented*.[32]]

only in HIDDENINFO*:*
[Player 2 only observes the result of the investment decision, but not whether player 1 has chosen Investment X or Investment Y. However, he has the possibility to change the investment by clicking the button "Decision Player 1" to find out if player 1 has chosen Investment X or Investment Y. The click costs 0.1 points. Player 2 can also make his distribution decision without informing himself about the decision of player 1.]

Payout

The payout of the two players from a round is calculated as follows:

- Player 1: Initial equipment - investment costs + payout from the project + payout from the distribution decision of player 2
- Player 2: Payout from the project + payout from own distribution decision

After player 2 has made his distribution decision, new groups of two are formed and a new independent round (consisting of an investment decision and a distribution

[31] We here chose to give as an example a split that could not be chosen by participants because they could only choose integer amounts. This was done on purpose to reduce potential anchoring effects from the example. This footnote was not included in the original instructions.

[32] The literal translation of the original instruction is "After he has made both decisions, it is resolved which decision player 1 has actually made. The actual decision determines which of the decisions is implemented by player 2". This sounds more ambiguous in English than in the original German text but could be misunderstood as subjects receiving feedback between rounds which was not the case. During the experiment, it was explained to participants that they would not receive feedback between rounds. For clarity, we changed the text to the part in italics here.

decision) begins. At the end of the experiment, one of the 4 rounds is randomly selected by drawing a card from a deck. Only this round from part 1 is relevant for payment.

For the first part of the experiment, an exchange rate of 0.2 from points to Euro applies, i.e. 10 points equal 2 Euro.

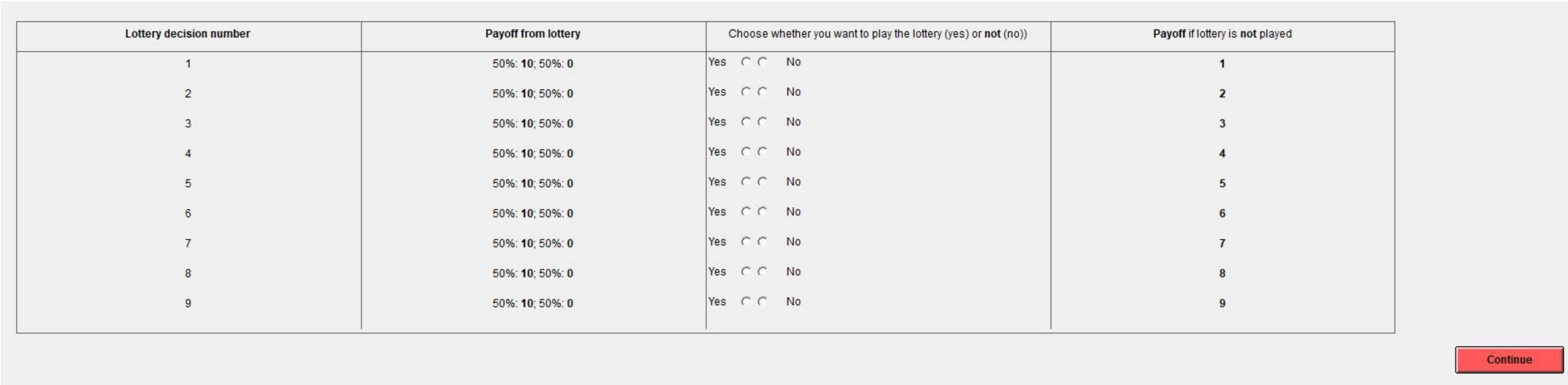
Your income from the experiment is calculated from your income from the first part of the experiment plus your income from the second part of the experiment plus a fixed payment of 5 Euro for participating.

In order to minimize the effort needed for payout at the end of the experiment, we round up your income from each part of the experiment to the next 10 cent amount.

The experiment is concluded with a questionnaire. Afterwards, each player will receive his payout privately and in cash.

Screenshots of the main decision screens

Figure 8: Risk preferences elicitation task

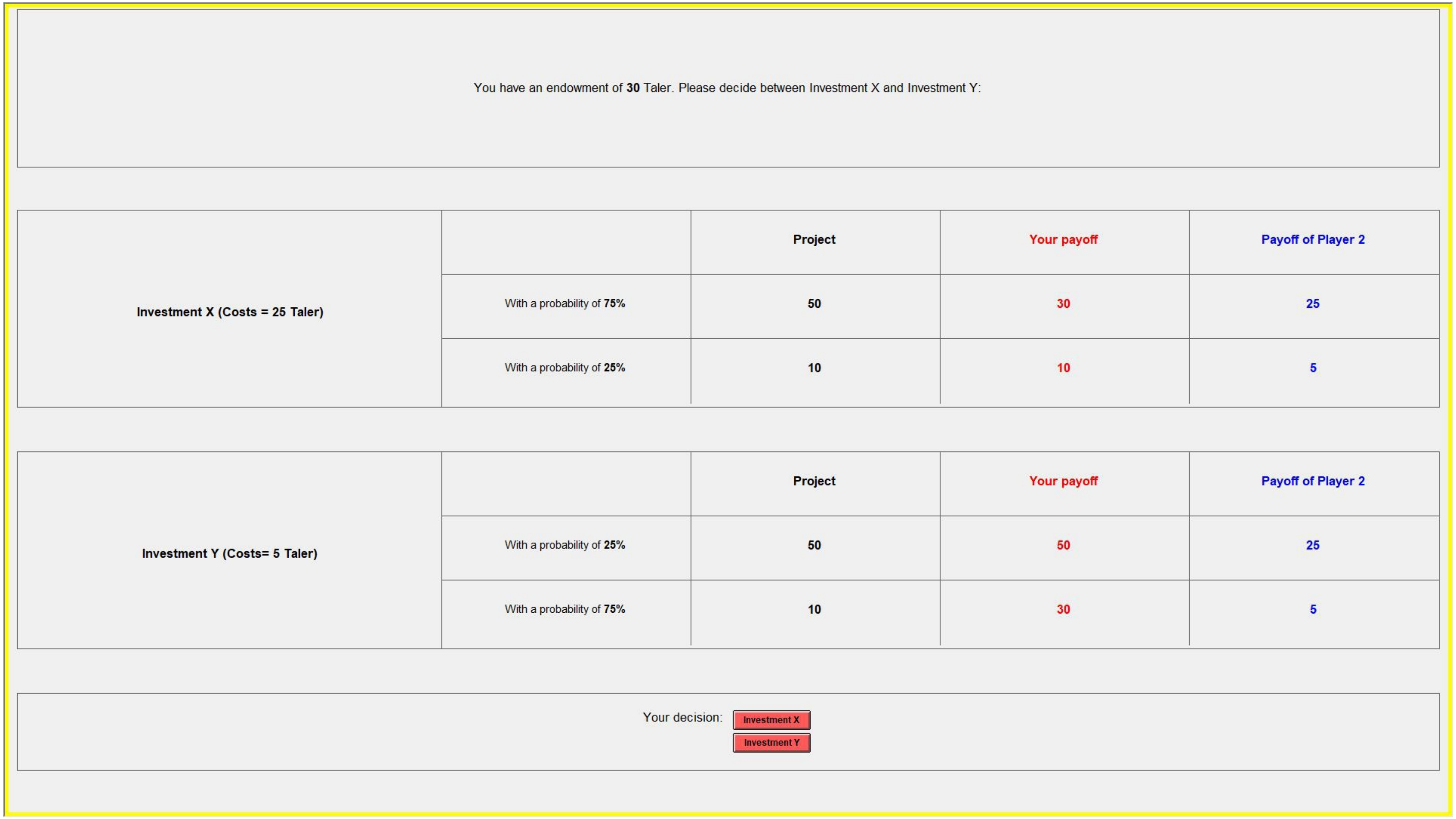


Lottery decision number	Payoff from lottery	Choose whether you want to play the lottery (yes) or **not** (no))	**Payoff** if lottery is **not** played
1	50%: **10**; 50%: **0**	Yes ◯ ◯ No	**1**
2	50%: **10**; 50%: **0**	Yes ◯ ◯ No	**2**
3	50%: **10**; 50%: **0**	Yes ◯ ◯ No	**3**
4	50%: **10**; 50%: **0**	Yes ◯ ◯ No	**4**
5	50%: **10**; 50%: **0**	Yes ◯ ◯ No	**5**
6	50%: **10**; 50%: **0**	Yes ◯ ◯ No	**6**
7	50%: **10**; 50%: **0**	Yes ◯ ◯ No	**7**
8	50%: **10**; 50%: **0**	Yes ◯ ◯ No	**8**
9	50%: **10**; 50%: **0**	Yes ◯ ◯ No	**9**

Continue

We used a choice-list design to elicit risk preferences. Subjects were confronted with a list of nine binary choices, out of which one was drawn to be payoff-relevant at the end of the experiment if the risk preferences were selected to be relevant for payment. In each choice, the subject had to decide between playing a lottery that delivered a payoff of either 10 or 0 points with a probability of 50payment. The secure payment varied from 1 point in the first binary choice to 9 points in the last binary choice. We use the first choice in which subjects chose the safe payment instead of playing the lottery as a measure for subjects' risk tolerance. If a subject chose the secure payment in the first decision, they were very risk averse, while they were risk seeking if they chose the lottery in the first eight decisions and switched to the safe payment only in the last row.

Figure 9: Player 1's main decision screen



You have an endowment of **30** Taler. Please decide between Investment X and Investment Y:

		Project	Your payoff	Payoff of Player 2
Investment X (Costs = 25 Taler)	With a probability of **75%**	**50**	**30**	**25**
	With a probability of **25%**	**10**	**10**	**5**

		Project	Your payoff	Payoff of Player 2
Investment Y (Costs= 5 Taler)	With a probability of **25%**	**50**	**50**	**25**
	With a probability of **75%**	**10**	**30**	**5**

Your decision: Investment X / Investment Y

Figure 10: Player 2's main decision screen FULLINFO

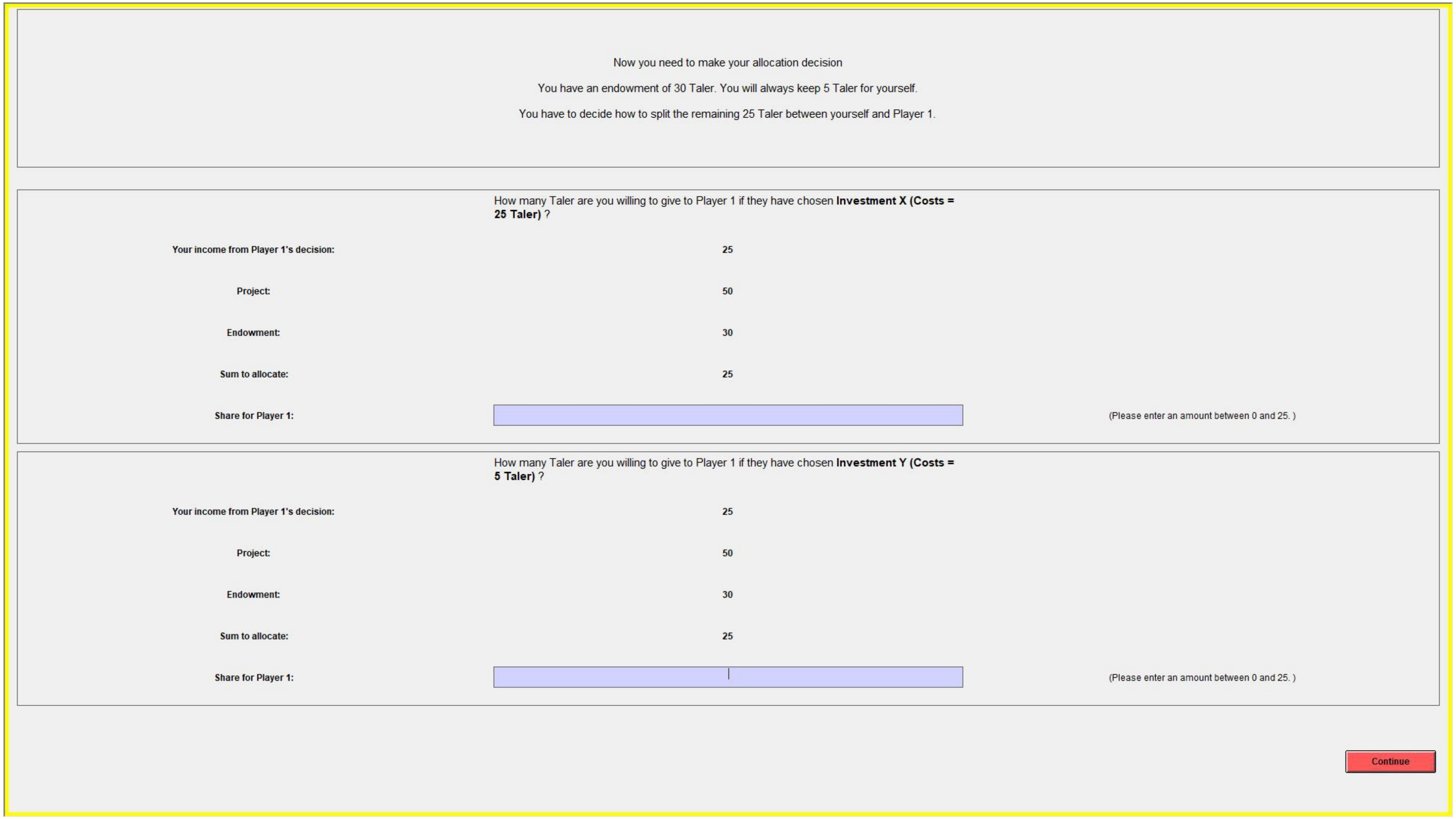


Figure 11: Player 2's main decision screen HIDDENINFO

Instructions for the second part of the experiment (displayed on screen)

In the following, you will make two decisions of which one will be payoff-relevant. Which one it is will be communicated at the end of the experiment.

Task 1

In this task, your decision will only affect your own payoff. The exchange rate from points to Euro is 0.2, i.e. 10 points correspond to 2 Euros.

Imagine there are 100 players of Type 1. In your opinion, how many players decided in favor of Investment X in the first round of the experiment?

If your estimate is correct, you will receive 15 points. If your estimate deviates from the correct number, you will lose 0.1 points per incorrectly estimated person.

Please decide now. If something is unclear, please raise your hand and we will come to you.

Task 2

In this task, your decision only affects your own payoff. Your payoff depends on your own decision and (potentially) a randomly drawn number. The exchange rate from points to Euro is 0.5, i.e. 2 points correspond to 1 Euro.

This task consists of a sequence of decisions to play or not to play a lottery. With a probability of 50% the lottery yields a payment of 0 points; with a probability of 50% it yields a payment of 10 points. If you decide against playing the lottery, you will receive a certain payment. This certain payment varies across the different decisions. In the first decision, it is 1 point, in the last decision, it is 9 points. For each decision, you can find the certain payment below.

If this task is chosen to be payoff-relevant, first a line will be determined randomly. Each line has the same probability of being chosen. Your decision for this line will implemented. If you have chosen the certain payment, you will receive it. If you have chosen the lottery, it will be played and you will receive 0 or 10 points, each with the same probability.

If something is unclear, please raise your hand and we will come to you.