University of Zurich

Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 125

Does Willful Ignorance Deflect Punishment? – An Experimental Study

Björn Bartling, Florian Engl and Roberto A. Weber

June 2013

Does Willful Ignorance Deflect Punishment? – An Experimental Study*

Björn Bartling
University of Zurich

Florian Engl
University of Zurich

Roberto A. Weber
University of Zurich

June 19, 2013

Abstract: This paper studies whether people can avoid punishment by remaining willfully ignorant about possible negative consequences of their actions for others. We employ a laboratory experiment, using modified dictator games in which a dictator can remain willfully ignorant about the payoff consequences of his decision for a receiver. A third party can punish the dictator after observing the dictator's decision and the resulting payoffs. On the one hand, willfully ignorant dictators are punished less if their actions lead to unfair outcomes than dictators who reveal the consequences before implementing the same outcome. On the other hand, willfully ignorant dictators are punished more than revealing dictators if their actions do not lead to unfair outcomes. We conclude that willful ignorance can circumvent blame when unfair outcomes result, but that the act of remaining willfully ignorant is itself punished, regardless of the outcome.

Keywords: Willful ignorance, third party punishment, dictator game, fairness

JEL classification: C91, D63

* We would like to thank Martin Dufwenberg, Bertil Tungodden, and participants of the ABEE Symposium 2012 on Behavioural Economics in Markets and Organizations in Amsterdam, the 2012 Zurich Workshop in Economics, the Sixth Annual NYU-CESS Conference on Experimental Political Science 2013, the 2013 Spring School in Behavioral and Experimental Economics in San Diego, and the 2013 Asia-Pacific ESA Conference in Tokyo for valuable comments.
Contact: Department of Economics, Blümlisalpstrasse 10, CH-8006 Zurich. Bartling: bjoern.bartling@econ.uzh.ch, Tel.: +41 44 634 3722; Engl: florian.engl@econ.uzh.ch, Tel.: +41 44 634 2006; Weber: roberto.weber@econ.uzh.ch, Tel.: +41 44 634 3688.

"A man is responsible for his ignorance." – Milan Kundera, *Laughable Loves*

1. Introduction

Many important decisions involve tradeoffs between personal benefits and impacts on the welfare of others. In such situations, there is often the possibility of remaining uninformed about how one's actions affect others. Such "willful ignorance" may provide a justification for self-interested behavior – after all, how can one be faulted for taking an action that one believes creates no harm for others? Thus, while a decision maker is typically held responsible for knowingly committing an action that hurts others, the attribution of responsibility is less clear when he acts without knowledge of consequences. Such reasoning may even hold when the decision to remain ignorant is made privately, as ignorance allows one to act selfishly without direct confrontation with the consequences for others or the associated guilt (Dana, Weber, and Kuang, 2007). Thus, strategically manipulating one's information about the consequences of one's actions for others provides a path through which ignorance, even when deliberate, might provide insulation from responsibility or blame.

In law, ignorance is not recognized as an excuse, as the legal principle "ignorantia juris non excusat" demonstrates. For example, when driving 65 mph, ignorance that the speed limit is 50 mph, fails to provide legal exculpation. Nevertheless, in corporate and political contexts, individuals often present ignorance as an excuse for why they should not be held responsible for adverse outcomes that they caused. For example, following corporate scandals and fraud, CEOs and board members often excuse their role by claiming they were not aware of what took place further down the hierarchy. Examples include former Enron CEO Kenneth Lay, who claimed ignorance about any accounting irregularities at the failed firm, and Rupert Murdoch, who was directly accused of showing "willful blindness" concerning the phone-hacking practices at News Corporation.[2] Following the 2008 financial crisis, many individuals and institutions involved in the sale of deceptively valued and marketed investment products deflected responsibility with the claim that these products were too difficult to understand.[3] In the political sphere, public officials

[2] See http://www.businessweek.com/stories/2006-02-05/commentary-ken-lays-audacious-ignorance and http://www.guardian.co.uk/media/2012/may/01/phone-hacking-report-wilful-blindness

[3] See http://www.nytimes.com/2009/03/12/business/12crime.html?pagewanted=1&_r=1&th&emc=th

sometimes argue that being unaware of acts that were committed in an administration should exonerate them from blame.[4]

Prior research in economics has demonstrated that decision makers seize upon strategies to act self-interestedly at the expense of others, when presented with opportunities for avoiding blame or responsibility.[5] An important but largely open question, however, is to what extent such strategies are, in fact, effective in deflecting blame.[6]

Our study directly addresses this issue, by attempting to quantify the extent to which engaging in willful ignorance allows a decision maker to deflect external blame for his actions and their consequences. To this end, we conduct a laboratory experiment in which some participants can choose to remain ignorant about the consequences of their actions on others and in which other participants have the opportunity to impose costly monetary punishments after observing behavior and the resulting outcomes. We interpret the assigned punishment points as a measure of blame and responsibility attribution for an action and its consequences.

More precisely, in our experiment a dictator plays a binary dictator game under one of two possible states of the world. The state of the world is chosen by a random device and determines whether an action that is personally beneficial for the dictator benefits or harms the receiver. The dictator can decide whether or not to learn the true state, and faces no cost for acquiring this information. The realized state is irrelevant for the dictator's payoffs, meaning that ignorance creates no uncertainty about the dictator's payoffs, but enables the dictator to remain ignorant about the effects of his action on others. Thus, our design affords the dictator the opportunity to remain willfully ignorant regarding the consequences of his actions for others.

Our focus is not on the effects of willful ignorance *per se*, however (cf. Dana, et al., 2007), but instead on the extent to which remaining willfully ignorant allows the dictator to avoid blame and responsibility when a bad outcome results for the recipient. Therefore, in our

[4] In fact, political science has long recognized the ability to avoid blame as an important determinant of a politician's success (Weaver, 1986) and ignorance as a potential strategy to do so (McGraw, 1991).

[5] Some research demonstrates that decision makers hide behind uncertainty – both their own and that of others – about what outcomes will result or how such outcomes were produced in order to keep more money in a distributional context (Dana et al., 2007; Andreoni and Bernheim, 2009; Ockenfels and Werner, 2012). In some cases, this can even mean that people are willing to accept less money in order to forgo the opportunity to share – and have the other person know that sharing could have taken place (Broberg, Ellingsen, and Johannesson 2007; Dana, Cain, and Dawes 2006; Lazear, Malmendier, and Weber 2012). Hamman, Loewenstein, and Weber (2010) show that delegating distributive decisions to others similarly provides a justification for self-interested behavior.

[6] Research into the effectiveness of blame-avoidance strategies has only recently started to gain traction. For example, Bartling and Fischbacher (2012) show experimentally that delegating a decision that leads to an unfair allocation is an effective way to shift blame from oneself toward the person to whom the decision is delegated.

experiment a third party observes the actions of the dictator and the outcome of the game and decides whether and to what extent to punish the dictator for his behavior.

Our results show that, when outcomes detrimental to the recipient result, ignorance is indeed effective in reducing punishment. That is, when taking an action that increases one's own welfare, but also results in harm to others, it is better to have avoided knowledge that harm would occur. Thus, ignorance can help avoid blame.

At the same time, however, we find that willful ignorance itself is evaluated negatively, regardless of the consequences. That is, choosing to forgo information concerning the recipient's payoffs and acting in a self-regarding way incites punishment, even when the resulting state of the world is one in which the dictator's self-interest is also beneficial for the recipient. Thus, the mere act of avoiding information about how one's decisions affect others provokes blame and punishment.

As a result of the above two counteracting effects of ignorance on punishment by third parties, in expectation, willful ignorance does not yield a higher payoff than knowingly acting selfishly. That is, while ignorance provides some blame avoidance when bad outcomes result, the fact that its use produces blame even when the outcomes are good makes it an ineffective strategy for obtaining higher payoffs.

However, the punishment *pattern* that is revealed in our study has important implications for how willful ignorance might interact with punishment outside the laboratory. Attention to the possibility of blame and punishment is often salient only when bad outcomes arise – e.g., following a scandal or harmful misdeed. The fact that decision makers are penalized less when acting under willful ignorance therefore suggests that willful ignorance may be a good strategy in contexts where punishment is unlikely to be considered absent some noticeably bad consequence. Thus, corporate and political leaders who suspect wrongdoing in the institutions they manage may, indeed, benefit from a strategy involving willful ignorance.

Our results are also important for economic theories of social preferences. We find significant differences in punishment for the same outcome, depending on whether the dictator revealed the state before making his choice. This cannot be explained by theories that incorporate social motives through preferences over final payoff distributions (Fehr and Schmidt 1999; Bolton and Ockenfels 2000), which predict the same punishment for an allocation, independently of the actions that led to the allocation. The qualitative comparative-static effect of willful ignorance on punishment is consistent with theories that incorporate intention-based reciprocity

as a motive (Rabin 1993; Dufwenberg and Kirchsteiger 2004; Falk and Fischbacher 2006; Sebald 2010). However, these models fail to predict our additional finding that willfully ignorant dictators are still punished less when the beneficial outcome occurs than when the unfair outcome obtains. That is, outcomes matter even for willfully ignorant dictators.

Research on procedural fairness recognizes that people care not only about distributions of final outcomes, but also about the procedures employed to implement outcomes (Frey, Benz, and Stutzer 2004; Bolton, Brandts, and Ockenfels 2005; Trautmann 2009; Krawczyk 2011; Fudenberg and Levine 2012). Our study contributes to this literature in that we show that punishment is not determined solely by consequences, but also by the process – the dictator's decision whether to acquire information – that leads to those consequences. Our research therefore also relates to other recent studies that find both *ex ante* fairness (equal opportunities, fair procedures) and *ex post* fairness (equal payoffs) to influence redistributive choices (Brock, Lange, and Ozbay 2013; Cappelen et al., 2013). We find that a simple model combining *ex ante* and *ex post* fairness, like the one suggested in Brock et al. (2013), is able to predict both the qualitative comparative-static effect of willful ignorance on the assigned punishment and the finding that punishment depends on consequences following willful ignorance.

There exists prior evidence that willful ignorance can be used to obtain more favorable wealth distributions, in the context of bilateral bargaining. Building on earlier experiments on bilateral bargaining with incomplete information about values, which demonstrated that more informed parties extract more favorable payoffs (Roth and Murnighan 1982), Kagel, Kim, and Moser (1996) show that responders in an ultimatum game are willing to accept very unequal monetary payoffs more often when the proposer is only partly informed about the receiver's payoffs than when the proposer has complete information. Thus, a party that is ignorant about the consequences of an offer for the other party can make less favorable offers. Conrads and Irlenbusch (2013) confirm that this applies to willful ignorance: offers to another party by a proposer in an ultimatum bargaining game who chooses to be ignorant are accepted more frequently than comparable offers by a fully informed proposer.

The remainder of the paper is organized as follows. Section 2 describes our experimental design. Section 3 summarizes our results with respect to the observed punishment pattern and the dictator's decisions. Section 4 discusses the predictions of different social preference models regarding the qualitative comparative-static effect of willful ignorance on punishment behavior. Finally, Section 5 concludes the paper.

2. Experimental Design

Our study uses one-shot binary dictator games that are modified to allow for willful ignorance and punishment. In the modified games, there are three players, as well as a move by nature that determines payoffs. Nature moves first, implementing one of two payoff states, ω_1 or ω_2, with equal probabilities, i.e., $p(\omega_1) = p(\omega_2) = 0.5$.

The state determines the relationship between a dictator's choices and the payoffs of a passive recipient, as depicted in Figure 1. More precisely, a dictator chooses between two options, a_1 and a_2. Regardless of the state, the dictator receives a payoff of 70 for choosing a_1 and 50 for choosing a_2. However, in ω_1, the recipient receives 10 for the dictator's choice of a_1 and 50 for a choice of a_2. In ω_2 the recipient's payoffs are reversed: 50 for a choice of a_1 and 10 for a_2.

Depending on the treatment, the dictator is either informed about the realized state or not. In a *hidden information* treatment, he is not initially informed, but he can choose whether to find out the state at no cost or remain willfully ignorant. The dictator then chooses between a_1 and a_2, either with or without knowledge of the state.

Finally, a third party can inflict punishment upon the dictator, after observing the dictator's choice, the realized state, and the resulting payoffs. Our primary interest is in these punishment decisions.

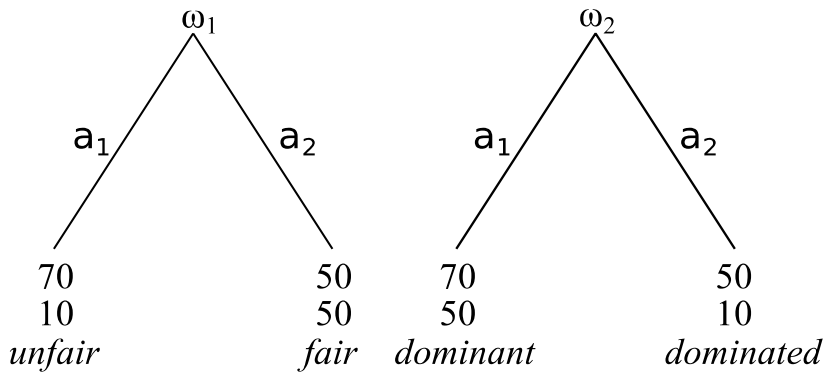


Figure 1: The dictator's choice options in state ω_1 and state ω_1. The dictator's monetary payoff is shown in the top row, the receiver's payoff in the bottom row.

Figure 1 also presents labels that provide an interpretation of the dictator's actions and their consequences, conditional on the realization of a particular state. In state ω_1, a choice of a_1

leads to an *unfair* allocation in that the dictator receives the highest possible payoff and the recipient the lowest one. Conversely, a choice of a_2 in state ω_1 leads to a *fair* allocation. Thus, in ω_1 there is a conflict between what is best for the dictator and for the recipient, as in standard dictator games. However, this conflict is entirely removed in state ω_2. Here a choice of a_1 is *dominant* for a dictator who cares both about her own payoff and that of the recipient, while a_2 leads to a *dominated* allocation of 50-10.[7] A conflict of interest thus exists in ω_1, but the dictator's and the receiver's payoffs are aligned in ω_2.

The third party has an endowment of 50 and can reduce the dictator's payoff after observing the dictator's allocation choice and the state of the world. Punishment is costly for the third party. For each unit of punishment allocated by the third party, the dictator's payoff decreases by 5 and the third party's payoff by 1. Punishment is constrained in that the dictator's payoff cannot be reduced below 10. Thus, for example, if the dictator's payoff is 70 before punishment, the third party can allocate any integer amount between 0 and 12 as punishment, to deduct up to 60 from the dictator's payoff. If the dictator's payoff is 50, the maximum punishment is 8 units, which decreases the dictator's payoff by 40.

Final payoffs are as follows. The dictator receives 70 or 50, depending on her choice of a_1 or a_2, minus five times the punishment points assigned by the third party. The receiver gets either 50 or 10, depending on the dictator's decision and the relevant state. The third party's payoff is 50 minus the units of punishment allocated to reduce the dictator's payoff.

We implemented two treatment variations that differ only with respect to the information that the dictator possesses regarding the state.

2.1 Baseline treatment

In the baseline treatment, it is common knowledge that the dictator is informed about the state of the world before he makes his decision between a_1 and a_2. Thus, the dictator is fully aware of whether the choice is between the *unfair* and *fair* allocations or the *dominant* and *dominated* ones. To elicit dictator's complete strategies, we implemented the strategy method. That is, we asked each dictator how he would decide if state ω_1 were realized and how he would decide if state ω_2 were realized. Only after the dictator made both choices, he learned the actual realized state, and he knew that his choice in this state would be binding.

[7] We use the labels "fair," "unfair," etc. for expositional reasons in the paper. In the experimental instructions, the dictator's choice options were neutrally framed as "Option 1" and "Option 2."

The third party was informed (i) about the state of the world and (ii) the dictator's choice in this realized state, and could then assign punishment points to decrease the dictator's payoff. We also applied the strategy method to elicit the punishment choices. That is, we asked the third party to indicate how many points she would deduct from the dictator's payoff for both possible choices by the dictator in both possible states of the world. Only after the third party made her decisions in all four possible cases, she learned the state of the world and the dictator's decision in this state. The third party knew that the chosen amount of punishment in the relevant case would be binding.

2.2 Hidden information treatment

In the hidden information treatment, it is common knowledge that the dictator is initially uninformed about the state of the world. Importantly, this uncertainty does not apply to the dictator's own payoffs, which are identical in both states. A choice of a_1 gives the dictator 70, while a_2 gives the dictator 50. Uncertainty only applies to the consequences of the two choices for the receiver's payoffs, as described in Figure 1. The dictator has the option to reveal the state before making his allocation decision. Following Dana et al. (2007), ignorance is the default, but revealing is costless and implemented by clicking a button on the decision screen.

If the dictator remains ignorant, his choice is between a payoff of 70 or 50 for himself. In this case, the dictator will never be informed about the underlying state of the world and he will thus never learn the receiver's payoff – i.e., whether the receiver got a payoff of 50 or 10. However, if the dictator reveals, he essentially places himself in the baseline condition. That is, he learns the state of the world and chooses either between the *unfair* and *fair* allocation in state ω_1, or between the *dominant* and *dominated* allocation in state ω_2.

As with the baseline treatment, we implemented the strategy method to elicit the allocation choices, where possible. That is, dictators first decided whether they wanted to acquire the payoff information or remain ignorant. If a dictator chose to remain ignorant, then he made a choice between a_1 and a_2, while if the dictator chose to acquire the payoff information, he then indicated choices of a_1 or a_2 for each of the two possible realized states. Only after the dictator made both choices, he learned the state of the world, and he knew that his choice in this state would be binding.

The third party was informed of (i) whether or not the dictator revealed the state, (ii) the realized state of the world, and (iii) what choice the dictator made, either in ignorance or

conditional on the realized state. The third party thus knew the state of the world even if the dictator chose to remain ignorant. The third party then decided how many punishment units to apply to the dictator's payoff. We again used the strategy method to elicit the punishment decisions by third parties for all possible states and actions by the dictator. Note that there are now eight possible cases, as all four possible allocations can result either after remaining ignorant or after revealing.

2.3 General procedures

Before the subjects entered the lab, they randomly drew a place card that specified at which computer terminal to sit. The terminal number determined both a subject's role and the group matching. After entering the lab, subjects found paper copies of the instructions at their assigned computer terminals. One third of the subjects read in the instructions that they were to be in the role of the dictator (neutrally labeled as "player A"). Two thirds of the subjects learned that they would be either in the role of the receiver ("player B") or in the role of the third party ("player C"). Subjects in these two roles both made choices as third parties; they learned of their actual roles only afterward. If they were assigned the role of the third party, then the chosen amount of punishment in the relevant case would be binding for their group. If they were assigned to the role of receiver, their decisions would have no impact on the group, as receivers made no decision in the game. This procedure enabled us to elicit punishment decisions, which are the focus of this paper, from two thirds of our subjects.

We conducted four sessions of the baseline treatment with 81 subjects in total (27 subjects in the role of the dictator and 54 subjects in the role of the receiver/third party). We also conducted four sessions of the hidden information treatment with 90 subjects in total (30 subjects in the role of the dictator and 60 subjects in the role of the receiver/third party).

All sessions took place at the decision laboratory of the Department of Economics at the University of Zurich in June 2012. The experiments were computerized with the software "z-Tree" (Fischbacher, 2007) and the recruitment was conducted with the software "ORSEE" (Greiner, 2003). Subjects were students from the University of Zurich and the Swiss Federal Institute of Technology (ETH) in Zurich. Students majoring in economics or psychology were not eligible to participate. Each subject participated in only one treatment. Subjects' instructions included comprehension questions that had to be answered correctly before the experiment could begin. A summary of the instructions was read aloud to ensure common knowledge of the

instructions. An English translation of the original German instructions for the hidden information treatment can be found in Appendix B. Baseline sessions lasted about 50 minutes, sessions with hidden information about 60 minutes. Payoffs from the game, denominated in "points" were converted into money at the rate of 2 points to CHF 1 (about $1 at the time of the experiment) at the end of the experiment. On average, subjects earned CHF 39.80 in the baseline sessions and CHF 41.30 in the hidden information sessions. These amounts include a show-up fee of CHF 15. The subjects received their payments privately to ensure the anonymity of the decisions taken during the experiment.

3. Results

3.1 Punishment Pattern

The focus of this paper is the pattern of third-party punishment for dictator allocation choices. Our particular interest is in studying how the dictator's choice to either remain ignorant or become informed about the receiver's payoffs influences punishment.

Figure 2 shows the average punishment that was assigned to the dictator for the different realized allocations in the baseline and in the hidden information treatment. The exact values can be read from Table 1. For instance, the left black bar in Figure 2 shows that the dictator is punished 19.72 points, on average, if he chooses the *unfair* allocation in state ω_1 in the baseline treatment.

In accordance with prior findings on third-party punishment (e.g., Fehr and Fischbacher 2004), the figure shows that the dictators are punished significantly more for knowingly implementing the *unfair* allocation than for the *fair* allocation. This holds true in the baseline and when the dictator chose to acquire the information in the hidden information treatment. When dictators remained ignorant, the difference in punishment for implementing the *unfair* vs. *fair* allocation was smaller, but also statistically significant. Thus, regardless of the dictator's knowledge or (willful) ignorance of the consequence to the receiver, a choice that results in an *unfair* allocation is punished more relative to one that results in a *fair* one ($p<0.01$ in all three comparisons, using a Wilcoxon signed rank test).[8]

[8] All tests reported in this paper are two-sided.

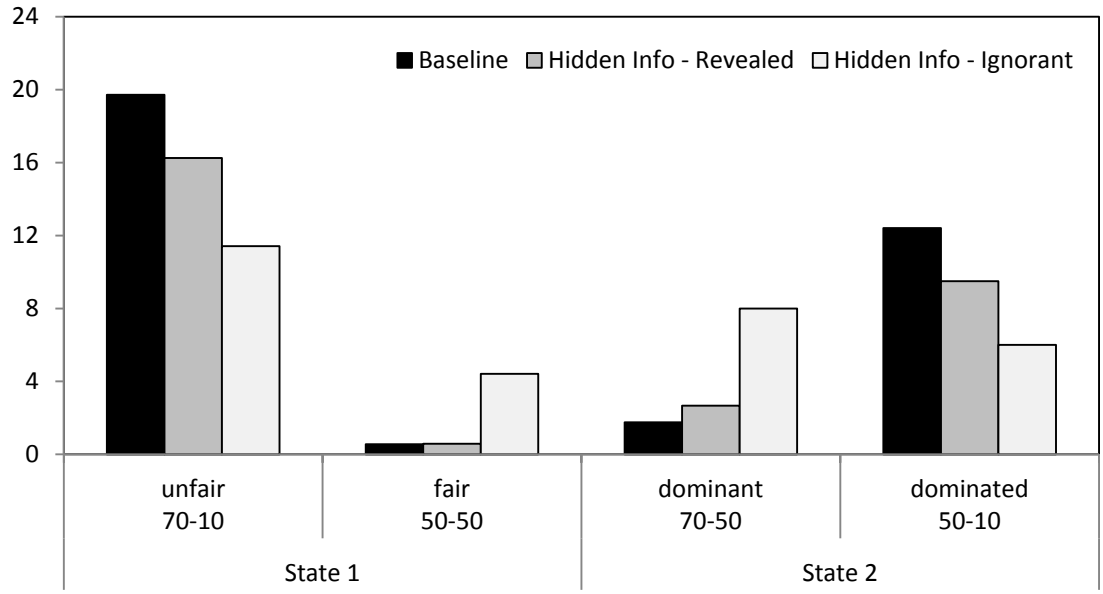


Figure 2: Average Punishment of the Dictator by the Third Party.

Our data show, however, that willful ignorance mitigates the punishment received by a dictator whose actions result in the *unfair* allocation. A willfully ignorant dictator who chooses 70 points for himself is punished significantly less if the *unfair* allocation realizes (11.42) compared to a dictator who directly chooses the *unfair* allocation when the consequences are known – i.e., after revealing (16.25) or in the baseline treatment (19.72) (Wilcoxon signed rank test, $p<0.01$, and Wilcoxon rank-sum test, $p=0.014$, respectively). Thus, our experiment reveals that willful ignorance can mitigate some of the blame and punishment received when knowingly implementing unfair outcomes.

> **Result 1**: *Willfully ignorant dictators are punished less for implementing an unfair outcome compared to dictators who knowingly chose the same outcome.*

However, the opposite pattern emerges when one considers what happens in cases where the resulting allocation is the *dominant* one, which is favorable to both the dictator and the receiver. Here, willfully ignorant dictators are punished significantly more (8.00) compared with dictators who choose the *dominant* allocation after revealing (2.76) or in the baseline treatment (1.76) (Wilcoxon signed rank test, $p=0.034$, and Wilcoxon rank-sum test $p=0.039$, respectively). Thus, willful ignorance itself appears to receive blame and punishment, even when it results in an outcome favorable to everyone.

Result 2: *Willfully ignorant dictators are punished more for implementing a dominant outcome compared to dictators who knowingly chose the same outcome. Willful ignorance is thus inherently blameworthy.*

Due to this opposing effect of willful ignorance on punishment, the difference in punishment between the *unfair* and the *dominant* allocation is much smaller when the dictator remained ignorant than when he revealed the state in the hidden information treatment or in the baseline treatment (3.42 vs. 13.58 and 17.96, respectively). Nevertheless, this difference is still highly significant (Wilcoxon signed rank tests, $p<0.01$ in all three cases).

Result 3: *Dictators, including willfully ignorant ones, are punished more if an unfair outcome obtains than if a dominant outcome obtains. Outcomes thus matter for punishment even under willful ignorance.*

We observe a similar pattern when a dictator chooses 50 for himself. In accordance with Result 1, if the choice is made under willful ignorance and the *dominated* allocation realizes, the dictator is punished significantly less (6.00) compared to a dictator who chooses *dominated* after revealing (9.50) or in the baseline treatment (12.41) (Wilcoxon signed rank test, $p=0.029$ and Wilcoxon rank-sum test, $p=0.011$, respectively). However, the willfully ignorant dictator is punished significantly more if the *fair* allocation realizes (4.42) compared to a dictator who chooses *fair* after revealing (0.58) or in the baseline treatment (0.56) (Wilcoxon signed rank test, $p<0.01$, and Wilcoxon rank-sum test, $p<0.01$, respectively).[9] This finding confirms Result 2. The difference in punishment between the *fair* and the *dominated* allocation is again much smaller when the dictator remained ignorant than when he revealed the state in the hidden information treatment or in the baseline treatment (1.59 vs. 8.92 and 11.85, respectively). This difference is at least marginally significant in all three cases (Wilcoxon signed rank tests, $p=0.052$, $p<0.01$ and $p<0.01$, respectively), which is consistent with Result 3.

To summarize, we find a consistent comparative-static effect of willful ignorance on punishment. On the one hand, whenever an allocation results in which the receiver receives the low payoff of 10, the dictator is punished significantly *less* when he remained ignorant than when he had the payoff information (Result 1). On the other hand, for allocations that are beneficial to

[9] As we report below, willfully ignorant dictators never chose 50 points for themselves. Also, none of the dictators who revealed chose *dominated* in state ω_2. In the baseline treatment, only one dictator chose *dominated*.

the receiver – i.e., when the receiver gets the high payoff of 50 – the dictator is punished significantly *more* when he remained ignorant (Result 2). Willful ignorance thus deflects blame and punishment for socially "bad" outcomes (the *unfair* or the *dominated* allocation). The fact that the dictator did not know for sure that the receiver would get a low payoff appears to serve as an acceptable excuse, to some extent. At the same time, willful ignorance is regarded as blameworthy in itself. A willfully ignorant dictator is punished significantly more than a dictator who reveals or a dictator in the baseline treatment when the receiver experiences no harm (in either the *fair* or the *dominant* allocation). Remaining ignorant means that the dictator risks the possibility of the receiver obtaining a low payoff, and this appears sufficient for inducing punishment by third parties. Finally, we observe that outcomes matter (Result 3). Dictators always receive more punishment when their actions yield the disadvantageous outcome for the receiver, regardless of the information possessed or acquired by the dictator.

Table 1: Average Punishment and Frequency of Punishment

		unfair (70-10)	*fair* (50-50)	*dominant* (70-50)	*dominated* (50-10)
Baseline	Average	19.72	0.56	1.76	12.41
	Frequency	0.61	0.04	0.13	0.50
Hidden Information – Revealed	Average	16.25	0.58	2.67	9.50
	Frequency	0.53	0.05	0.13	0.37
Hidden Information – Ignorant	Average	11.42	4.42	8.00	6.00
	Frequency	0.38	0.20	0.27	0.28

Finally, we do not find that revealing the state is treated differently from exogenously knowing the state. A comparison of the punishment for a dictator who reveals in the hidden information treatment with the punishment in the baseline treatment, where the dictator knows the state of the world by default, reveals no significant differences (Wilcoxon rank-sum tests, p=0.331, p=0.743, p=0.900, and p=0.196, for *unfair*, *fair*, *dominant*, and *dominated*, respectively).

A similar pattern to the one that we observe in punishment levels also emerges when we look at the comparative-static effect of willful ignorance on the frequency of punishment, presented in Table 1. A willfully ignorant dictator who chooses a_1 and a payoff of 70 for himself is punished less often if the *unfair* allocation results (38 percent), compared to a dictator who reveals (53 percent) or to the baseline treatment (61 percent) (McNemar test, p=0.012, and

Fischer exact test, p=0.024, respectively).[10] Conversely, if the *dominant* allocation results, a willfully ignorant dictator is punished more frequently (27 percent versus 13 percent, in both cases) (McNemar test, p=0.039, and Fischer exact test, p=0.101, respectively). Similarly, a willfully ignorant dictator who chooses 50 for himself is punished more often if the fair allocation results and less often if the dominated allocation results, compared to a dictator who reveals or to the baseline treatment, though the difference is not significant in all cases (McNemar tests, p=0.012 and p=0.180, and Fisher exact tests, p=0.010 and p=0.021, respectively).[11]

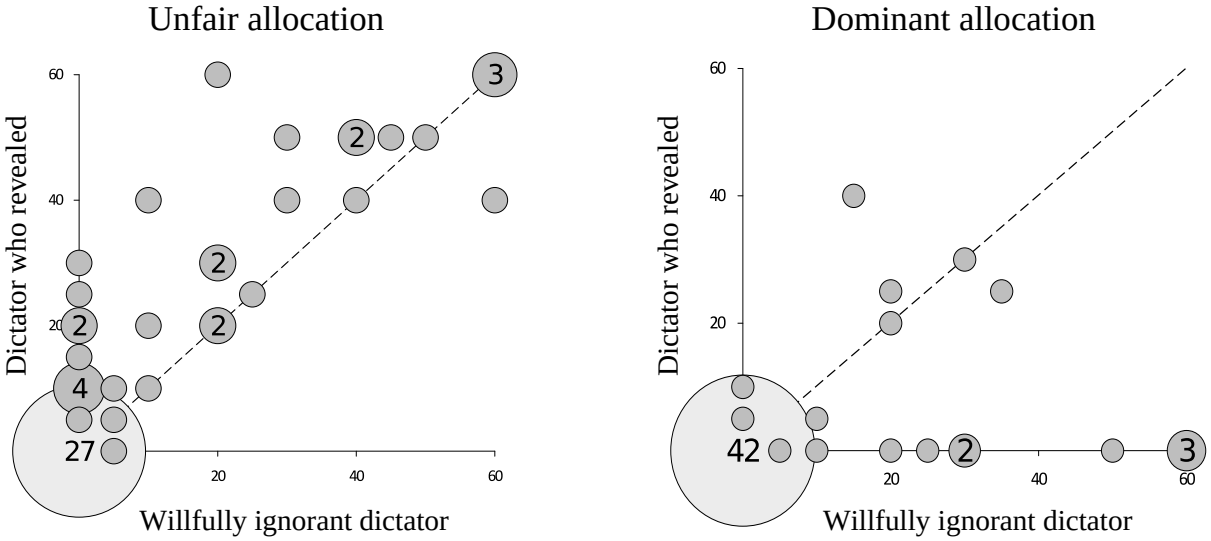


Figure 3: Individual Third Party's Punishment Assignment for the *Unfair* and *Dominant* Allocation Depending on the Dictator's Choice to either Reveal or Remain Ignorant.

Results 1 and 2 are further illustrated in Figure 3. The figure shows the individual third parties' punishment assignments in the hidden information treatment when either the *unfair* allocation (left panel) or the *dominant* allocation (right panel) are realized. Circles above (below) the 45-degree line indicate third parties who punished a dictator who first revealed more (less) than a willfully ignorant dictator. The numbers in the circles indicate the number of observations; circles without numbers represent one observation. For instance, when the *unfair* allocation realized, 27 third parties punished neither a willfully ignorant dictator nor a dictator who first

[10] Our results on the punishment level and frequency are confirmed by regression analyses that we conduct to complement the non-parametric tests. See Appendix A for the details.

[11] Consistent with our observation on levels of punishment, there is no difference in the frequency of punishment between the baseline and the hidden information treatments when the dictator reveals the payoff information (Fisher exact tests, p=0.451, p=1, p=1, p=0.186 for unfair, fair, dominant, and dominated, respectively).

revealed. Providing further support for the punishment pattern we observed earlier, of those third parties who did punish the *unfair* allocation, the majority assigned greater punishment to a dictator who first revealed than to a willfully ignorant dictator. The pattern is reversed when the *dominant* allocation realizes: the majority of those third parties who punished assigned more punishment to a willfully ignorant dictator than to a dictator who first revealed.

3.2 Expected Payoffs of Dictators

We now turn to the dictators' expected payoffs for each possible allocation choice strategy. There are four choice strategies in the baseline treatment, based on the two possible realized states and the two possible actions in each state. Because there is no uncertainty, these strategies are identified by the resulting outcomes: {*unfair*, *dominant*}, {*fair*, *dominant*}, {*unfair*, *dominated*}, and {*fair*, *dominated*}. In the hidden information treatment, the dictator can choose to either reveal the payoff information – in which case the same four strategies as in the baseline become available – or to remain willfully ignorant, in which case the two unconditional action choices, a_1 or a_2, are available. Table 2 shows the dictators' average expected payoffs, based on the punishment behavior of third parties, for each of the possible strategies in a treatment.

Table 2: Expected Payoffs of Dictators under Different Strategies

		Baseline	Hidden Information	
			Reveal	Ignorance
{*unfair, dominant*}	$(a_1 \mid \omega_1)\,(a_1 \mid \omega_2)$	59.26	60.54	
{*fair, dominant*}	$(a_2 \mid \omega_1)\,(a_1 \mid \omega_2)$	58.84	58.38	
{*unfair, dominated*}	$(a_1 \mid \omega_1)\,(a_2 \mid \omega_2)$	43.94	47.13	
{*fair, dominated*}	$(a_2 \mid \omega_1)\,(a_2 \mid \omega_2)$	43.52	44.96	
{*unfair / dominant*}	(a_1)	-		60.29
{*fair / dominated*}	(a_2)	-		44.79

Table 2 reveals, not surprisingly, that any strategy that pursues the *dominated* allocation in state ω_2 yields lower expected payoffs than strategies pursuing the *dominant* allocation (Wilcoxon signed rank tests for all five comparisons, $p<0.01$).[12] The three strategies that implement the *dominated* allocation yield similar expected payoffs.

[12] For each strategy, a distribution of payoffs is generated by the punishment behavior of the third parties, which we use as the basis of all tests in this subsection.

Our main interest is the effect of the dictator's choice to remain ignorant on his expected payoff. We first compare the strategies that select the same allocations. In this regard, there is little difference between the expected payoffs of a dictator who chooses to remain ignorant and selects action a_1 (60.29) and a dictator who reveals the payoff information, selects action a_1 regardless of the realized state (i.e., pursues the {*unfair, dominant*} allocations) and receives an expected payoff of 60.54 (Wilcoxon signed rank test, p=0.137). However, the difference in expected earnings of a willfully ignorant dictator who selects a_1 and a dictator who chooses the same action unconditionally in the baseline is marginally significant (60.29 vs. 59.26; Wilcoxon rank-sum test, p=0.075).[13]

We can also compare the strategy of remaining ignorant and selecting a_1 to revealing and acting fairly by giving the receiver a payoff of 50, regardless of the state – i.e., implementing the *fair* or *dominant* allocations – in the hidden information treatment and to acting fairly in the baseline treatment.[14] While the differences are small, the expected payoff of remaining ignorant and playing a_1 (60.29) is significantly higher than the expected payoff of either of these two other strategies (58.38 and 58.84; Wilcoxon signed rank test, p<0.01, and Wilcoxon rank-sum test, p<0.01, respectively).

The observation of very small payoff differences reflects our finding that willful ignorance has two countervailing effects on punishment, described in Results 1 and 2. When dictators forgo acquiring the payoff information and select action a_1, they are punished less if an *unfair* allocation results but more if a *dominant* allocation results. Thus, while a strategy relying on willful ignorance provides some diversion of punishment when unfair outcomes result, such a strategy also yields more punishment when fortune produces a favorable outcome for the receiver. As a result, willful ignorance provides similar expected payoffs to acting with information about payoffs.

3.3 Dictators' Strategies and Resulting Allocations

Finally, we consider the dictators' information acquisition decisions in the treatment with hidden information, as well as their allocation choices in all treatments.

[13] Comparing the expected payoffs in the baseline and the hidden information treatment when the dictator reveals, we find no significant differences in any of the four strategies; Wilcoxon rank-sum tests, p>0.1). This is in line with the observation that there are no significant differences in the punishment pattern.

[14] As we will see (in Section 3.3), these are the most frequently chosen strategies.

In the baseline treatment, 33 percent of dictators (9 out of 27) chose the action a_1 regardless of the state, which corresponds to the allocations {*unfair*, *dominant*}. Almost twice as many, or 63 percent (17 out of 27), chose the strategy that gave the receiver a payoff of 50 in either stage – e.g., a_2 in state ω_1 and a_1 in state ω_2, or {*fair*, *dominant*}. One subject chose action a_2 in state ω_2, implementing {*fair*, *dominated*}. This overall pattern of behavior is in line with earlier results on dictator games with punishment.[15]

In the hidden information treatment, 43 percent of the dictators (13 out of 30) remained ignorant about the consequences of their decision for the receiver. This percentage almost exactly matches the 44 percent of dictators who remained ignorant in Dana et al. (2007). All of the dictators who remained ignorant chose action a_1 {*unfair*, *dominant*}. Of those dictators who revealed the state, 12 percent (2 out of 17) choose a_1 unconditionally {*unfair*, *dominant*} and 88 percent (15 out of 17) choose a_2 in state ω_1 and a_1 in state ω_2 {*fair*, *dominant*}. Dictators who revealed the state thus chose the *fair* allocation in state ω_1 in the large majority of the cases. This shows that most dictators who reveal the state do so in order to be able to condition their allocation choice on the state of the world.

The dictators' strategies resulted in different frequencies of the possible allocations in the two treatments. In the baseline treatment, when state ω_1 realized, 33 percent of dictators (9 out of 27) chose the *unfair* allocation. The *unfair* allocation resulted with higher frequency (50 percent, or 15 of 30) in the hidden information treatment. In state ω_2, the *dominant* allocation resulted with almost the same frequency in the hidden information treatment (30 out of 30 cases) and the baseline treatment (26 out of 27 cases).

The fact that *unfair* allocations result more frequently under hidden information than in the baseline treatment resembles the findings in Dana et al. (2007). In their experiment the frequency of the *unfair* allocation increased from 26 to 63 percent. The interpretation of Dana, et al., is that the possibility to remain ignorant gives subjects the moral "wiggle room" to behave self-interestedly. While similar in direction, the effect in our experiment is much smaller and not statistically significant (Fisher exact test, p=0.284). Of course, a key difference between the two experiments is the presence of a punishment stage in our design. The threat of punishment alone potentially limits the extent to which subjects are willing to act as if willful ignorance absolves

[15] In Bartling and Fischbacher (2012), for instance, 63 percent of dictators selected a fair allocation in a binary dictator game with punishment that is comparable to our game if state ω_1 prevails.

them of responsibility. As we see, third parties still hold dictators responsible for their ignorance.[16]

4. Discussion of Behavioral Model Predictions

In this paper, we ask the empirical question whether willful ignorance can reduce punishment for a dictator who implements an unfair allocation. Our goal was not to design an experiment in order to distinguish between different behavioral models of punishment and social preferences. However, it is nevertheless instructive to discuss the qualitative comparative-static predictions of some prominent models regarding the impact of the dictator's choice to remain willfully ignorant on the punishment by the third party. Therefore, we now consider some of the leading models in the literature, and compare their qualitative predictions to the pattern of punishment behavior in our experiment.

Note first that the canonical model of pure self-interest predicts that no subject in the role of a third party will punish because punishment is costly. This prediction is clearly inconsistent with the data.

4.1 Outcome-Based Models of Social Preferences

Outcome-based models of social preferences (Fehr and Schmidt 1999; Bolton and Ockenfels 2000) assume that some people dislike payoff inequalities. Inequality-averse people might thus be willing to incur costs to reduce the payoff inequality – e.g., by punishing an unfair dictator. Importantly, these models are based on final monetary payoffs only and do not take into account how these allocations came about.

In the spirit of these models, we assume that the third party cares about the payoff difference between the dictator and the receiver. Consistent with this model class, we observe higher levels and frequencies of punishment for allocations with higher inequality – even for willfully ignorant dictators (Result 3). The important observation for our purposes, however, is that for a given allocation, purely outcome-based social preferences models do not predict a difference based on how that allocation was produced. Therefore, punishment should be the

[16] Moreover, while in Dana, et al., subjects who remained willfully ignorant never found out about the consequences for the receiver, dictators in our experiment received a "punishment signal" about the realized state of the world, due to the fact that third parties punished differently when the *unfair* allocation resulted than when the result was the *dominant* allocation. Thus, dictators lose some of the benefit of remaining ignorance, due to the information conveyed by punishment.

same, for example, when the *unfair* allocation is realized, for dictators who remained ignorant in the hidden information treatment, who revealed the state in the hidden information treatment, or who made choices in the baseline treatment. The data (Results 1 and 2) do not support this prediction.

4.2 Intention-Based Models of Social Preferences

A second class of social preference models focuses on the intentions behind an action (Rabin 1993; Dufwenberg and Kirchsteiger 2004; Seebald 2010). These models assume that people are willing to incur costs to punish unkind actions and to reward kind actions. The main difference with purely outcome-based models is that players respond to the perceived kindness of other players and not to the inequality of the resulting outcomes. The kindness of a player is measured as the distance from an "equitable" payoff, which is defined as the average between the highest and the lowest efficient payoff that a player can grant another player. The action of a player is perceived as kind (unkind) if he gives more (less) than the "equitable" payoff. Sebald (2010) extends the dynamic framework of Dufwenberg and Kirchsteiger (2004) to allow for games in which monetary payoffs also depend on chance. In this model, remaining ignorant can be interpreted as choosing a lottery over the receiver's payoff. We assume that the dictator's behavior towards the receiver determines the third party's assessment of the dictator's kindness, and that the third party derives utility from reciprocating the (un)kindness of the dictator.

Consider first the dictator's choice of the *unfair* allocation (70-10) in state ω_1 in the baseline treatment or after revealing in the hidden information treatment. The implementation of the *unfair* allocation is evaluated as unkind because it leaves the receiver with less than the equitable payoff of 30 (the average of the receiver's highest and lowest possible payoff of 50 and 10, respectively). Second, the implementation of the *fair* allocation (50-50) in state ω_1 in the baseline treatment or after revealing in the hidden information treatment is evaluated as kind because it leaves the receiver with more than the equitable payoff of 30. Finally, remaining willfully ignorant leads to the lottery $0.5 \circ 50 + 0.5 \circ 10$ over the receiver's payoff, regardless of whether the dictator chooses a_1 or a_2. Since the expected payoff of this lottery equals the equitable payoff of 30, remaining ignorant is evaluated as neither kind nor unkind.

Thus, if a third party reciprocates a dictator's unkindness with punishment, these models correctly predict Results 1 and 2; i.e., more (less) punishment for a dictator who implements an *unfair* (*fair*) allocation after revealing the state, relative to a dictator who remained willfully

ignorant.[17] The same prediction pattern prevails in state ω_2, where this class of models correctly predicts why a dictator who revealed the state is punished more (less) when the *dominated* (*dominant*) allocation results, compared to a dictator who remained willfully ignorant.[18]

However, intention-based theories cannot explain why a willfully ignorant dictator is punished differently depending on which state results (Result 3). After the decision to remain ignorant, the final allocation does not influence the third party's evaluation of the dictator's kindness and should thus not affect the punishment decision.

4.3 Hybrid Models of Outcome- and Intention-Based Social Preferences

Two factors determine the kindness of an action in Falk and Fischbacher's (2006) theory of reciprocity. First, the kindness of a player – such as the dictator – is determined by the expected payoff inequality, where the expectation is taken at the dictator's decision node. Second, it is decisive whether an action is intentional or not, which depends on the dictator's choice options. The dictator is considered as unkind (kind) if he implements an allocation that favors him (the receiver) *in expectation*, and he is considered as even more unkind (kind) if his choice is intentional. We assume that the dictator's choice options and the expected payoff difference between dictator and receiver determine the third party's evaluation of the dictator's kindness, and that the third party derives utility from reciprocating the (un)kindness of the dictator.

Consider first the dictator's choice of the *unfair* allocation (70-10) in state ω_1 in the baseline treatment or after revealing in the hidden information treatment. The implementation of the *unfair* allocation is intentional because the *fair* allocation (50-50) could have been chosen, and a (expected) payoff difference of 60 results. Remaining ignorant grants the recipient an expected payoff of 30. Again, the *fair* allocation could have been chosen (after revealing) and, hence, the action is intentional. In both cases, therefore, the dictator's action is intentional but the expected payoff difference is lower in case of a willfully ignorant dictator. The model thus makes the correct qualitative comparative-static prediction that a willfully ignorant dictator is punished less if the *unfair* allocation is implemented compared to a dictator who reveals the state or a dictator in the baseline treatment (Result 1).

[17] Note however, that, given the choice options of the third party, the model predicts positive punishment only for the implementation of an unkind act and no punishment for the implementation of a neutral or kind act. This is because the third party cannot reciprocate a kind act by the dictator with a reward.

[18] The choice of the *dominated* allocation in the baseline treatment or after revealing in the hidden information treatment is unkind, while the choice of the *dominant* allocation is kind. A willfully ignorant dictator, again, is neither kind nor unkind.

Consider now the dictator's choice of the *fair* allocation (50-50) in state ω_1 in the baseline treatment or after revealing in the hidden information treatment. Since an equitable payoff is implemented, the dictator's action is not perceived as unkind and the prediction is no punishment. If the dictator remained ignorant, however, the expected payoff difference is 20. Moreover, the action is also intentional because the dictator could have revealed the state and chosen (50-50). It follows that the qualitative comparative-static prediction of the model is also in line with the second effect of willful ignorance on the punishment pattern, namely that a willfully ignorant dictator is punished more if the *fair* allocation results (Result 2). The same prediction pattern prevails in state ω_2.[19]

Again, while this model correctly predicts our first two main results, it cannot explain why a willfully ignorant dictator is punished differently depending on which state results (Result 3), because the dictator's kindness is evaluated at his decision node and not after the revelation of uncertainty. As was the case in purely intention-based models, the third party's punishment decision should thus not be influenced by the resulting state.

4.4 Models of procedural fairness

Models of procedural fairness assume people do not only care about outcomes but also about the procedures that lead to these outcomes (Frey, Benz, and Stutzer 2004; Bolton et al. 2005; Trautmann 2009; Krawczyk 2011; Fudenberg and Levine 2012; Brock et al. 2013; Cappelen et al., 2013). An important example of such a procedure is the notion of "equal opportunities." One interpretation of this notion is that not only *ex post* realized payoff distributions are important but also *ex ante* expected payoff distributions. Brock et al. (2013), for example, extend the Fehr and Schmidt (1999) model to allow for a convex combination of *ex ante* and *ex post* payoff comparisons. Without uncertainty, *ex ante* and *ex post* fairness coincide, but with uncertainty a player's fairness evaluations depend on the weight that he places on *ex ante* and *ex post* payoff

[19] If the *dominated* allocation (50-10) is chosen, the choice is intentional in all cases because the *dominant* allocation was available. Since the (expected) payoff difference is higher in case of a dictator in the baseline treatment or after revealing in the hidden information treatment, the model correctly predicts lower punishment for a willfully ignorant dictator. If the *dominant* allocation (70-50) is chosen in the baseline treatment or after revealing in the hidden information treatment, first, the (expected) payoff difference is 20 and, second, the choice is not intentional because the only alternative gives the receiver an even lower payoff. If the *dominant* allocation however results after remaining ignorant, there are two differences. First, the expected payoff difference is higher (40). Second, the choice is even intentional because the dictator could have revealed the state and chosen an allocation that gives the receiver a payoff of 50 and not only the expected payoff of 30. Hence, the model again correctly predicts higher punishment for a willfully ignorant dictator.

differences. We assume that the third party cares about the payoff difference between the dictator and the receiver and, for simplicity, we also initially assume that the third party places weight *only* on *ex ante* payoff differences.

If the *unfair* allocation (70-10) results in state ω_1 in the baseline treatment or after revealing in the hidden information treatment, the (*ex ante*) payoff difference is 60. Since the *ex ante* payoff difference in case of a willfully ignorant dictator who chose a_1 is only 40, the model makes the correct qualitative comparative-static prediction that a willfully ignorant dictator is punished less (Result 1). If the *fair* allocation (50-50) results in state ω_1 in the baseline treatment or after revealing in the hidden information treatment, the (*ex ante*) payoff difference is zero. Since the *ex ante* payoff difference is 20 in case of a willfully ignorant dictator who chose a_2, the model again makes the correct qualitative comparative-static prediction that a willfully ignorant dictator is punished more (Result 2). The same qualitative prediction pattern prevails in state ω_2.

A simple model that places sufficient weight on the *ex ante* payoff differences is thus able to qualitatively predict the comparative-static effect of the dictator's choice to either remain ignorant or to reveal the state of the world on the assigned punishment. If the model additionally places some weight on *ex post* payoff differences, it also correctly predicts that a willfully ignorant dictator who chose a_1 (a_2) is punished more if the *unfair* (*dominated*) allocation results than if the *dominant* (*fair*) allocation results (Result 3).

5. Conclusion

This paper studies costly third-party punishment. Consistent with prior research, we find that individuals who observe others behaving unfairly often incur personal costs to punish such unfairness.

Our focus in this study is on how the opportunity to remain willfully ignorant – by avoiding information on the consequences of one's actions for others – affects the extent to which individuals are held accountable and punished by third parties for the consequences of their actions. Discussions of responsibility in political and corporate scandals are often accompanied by claims of ignorance that could have been resolved if the involved parties had sought out the relevant information. It is important, therefore, to understand whether such strategies are effective for deflecting blame and punishment.

Our evidence in this regard reveals an interesting pattern. When bad consequences result – i.e., *unfair* or *dominated* allocations – a decision maker is punished less when he chose not to know the consequences of his actions than when he either first chose to find out or was informed by default. Thus, when decision makers would be blamed or punished for something bad happening, they can avoid some of this blame or punishment through (the perception of) ignorance, even when deliberately chosen.

On the other hand, we also observe that the act of remaining willfully ignorant is, itself, punished. That is, when good outcomes result following a decision maker's choice to remain ignorant – i.e., *fair* or *dominant* allocations – a decision maker is nevertheless punished to some extent. Thus, while willful ignorance may be effective for deflecting blame when negative consequences are likely, awareness that one is using willful ignorance as part of a strategy is perceived negatively and inherently punishable.

In short, by remaining willfully ignorant, decision makers deflect some punishment when bad consequences arise, due to the fact that something good could have happened. Conversely, when good outcomes result from decisions made under willful ignorance, the fact that less desirable outcomes could have obtained provides grounds for punishment. Nevertheless, even under willful ignorance, punishment is still higher when bad consequences arise than when good outcomes result. Such punishment behavior by third parties is consistent with simple behavioral social preference models that combine *ex ante* and *ex post* fairness concerns (Brock et al. 2013; Cappelen et al. 2013).

For dictators in our experiment, willful ignorance is not a better strategy in expectation than acquiring payoff information. This is mainly because the third parties punish willful ignorance even when fortune produces a favorable outcome for the recipient.

However, this punishment *pattern* may have very different consequences outside the laboratory. Since outside the lab attention to the possibility of punishing someone is often salient only when bad outcomes arise, the fact that decision makers receive lighter sanctions when willfully ignorant suggests this may be a good strategy in contexts where punishment is unlikely to be considered absent some noticeably bad consequence. Thus, for example, the symmetry of punishment possibilities in our experiment may be an important factor not present in natural contexts, where punishment may be considered only in cases where bad outcomes have indeed been produced. In such situations, our results suggest that, indeed, willful ignorance may circumvent blame and punishment.

References

Andreoni, James, and B. Douglas Bernheim. 2009. "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects." *Econometrica* 77(5): 1607–36

Bartling, Björn, and Urs Fischbacher. 2012. "Shifting the Blame: On Delegation and Responsibility." *Review of Economic Studies* 79(1): 67–87.

Bolton, Gary E, Jordi Brandts, and Axel Ockenfels. 2005. "Fair Procedures: Evidence from Games Involving Lotteries." *The Economic Journal* 115(506): 1054–76.

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

Broberg, Tomas, Tore Ellingsen, and Magnus Johannesson. 2007. "Is generosity involuntary?" *Economics Letters* 94(1): 32–37.

Brock, J. Michelle, Andreas Lange, and Erkut Y. Ozbay. 2013. "Dictating the Risk – Experimental Evidence on Giving in Risky Environments." *American Economic Review* 103(1): 415-37.

Cappelen, Alexander W., James Konow, Erik Ø Sørensen, and Bertil Tungodden. 2013. "Just Luck: An Experimental Study of Risk Taking and Fairness." *American Economic Review* 103(4): 1398-1413.

Conrads, Julian, and Bernd Irlenbusch. 2013. "Strategic Ignorance in Bargaining." *Journal of Economic Behavior and Organization*, forthcoming.

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

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

Dufwenberg, Martin, and Georg Kirchsteiger. 2004. "A theory of sequential reciprocity." *Games and Economic Behavior* 47(2): 268–98.

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

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

Fehr, Ernst, and Urs Fischbacher. 2004. "Third-party punishment and social norms." *Evolution and Human Behavior* 25(2): 63–87.

Fischbacher, Urs. 2007. “z-Tree: Zurich toolbox for ready-made economic experiments.” *Experimental Economics* 10(2): 171–78.

Frey, Bruno S., Matthias Benz, and Alois Stutzer. 2004. “Introducing Procedural Utility: Not Only What, but Also How Matters.” *Journal of Institutional and Theoretical Economics* 160: 377-401.

Fudenberg, Drew, and David K. Levine. 2012. “Fairness, risk preferences and independence: Impossibility theorems.” *Journal of Economic Behavior & Organization* 81(2): 606–12.

Greiner, Ben. 2003. “An Online Recruitment System for Economic Experiments.” In *Forschung und wissenschaftliches Rechnen 2003. GWD Bericht 62*, eds. Kurt Kremer and Volker Macho. Göttingen: Ges. für Wiss. Datenverarbeitung, p. 79–93.

Hamman, John R, George Loewenstein, and Roberto A Weber. 2010. “Self-Interest through Delegation: An Additional Rationale for the Principal-Agent Relationship.” *American Economic Review* 100(4): 1826–46.

Kagel, John H., Chung Kim, and Donald Moser. 1996. “Fairness in Ultimatum Games with Asymmetric Information and Asymmetric Payoffs.” *Games and Economic Behavior* 13(1): 100-10.

Krawczyk, Michal. 2011. “A model of procedural and distributive fairness.” *Theory and Decision* 70(1): 111–28.

Lazear, Edward P., Ulrike Malmendier, and Roberto A. Weber. 2012. “Sorting in Experiments with Application to Social Preferences.” *American Economic Journal: Applied Economics* 4(1): 136–63.

McGraw, Kathleen M. 1991. “Managing Blame: An Experimental Test of the Effects of Political Accounts.” *The American Political Science Review* 85(4): 1133–57.

Ockenfels, Axel, and Peter Werner. 2012. “‘Hiding behind a small cake’ in a newspaper dictator game.” *Journal of Economic Behavior & Organization* 82(1): 82–85.

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

Roth, Alvin E., and J. Keith Murnighan. 1982. “The Role of Information in Bargaining: An Experimental Study.” *Econometrica* 50(5): 1123–42.

Sebald, Alexander. 2010. “Attribution and reciprocity.” *Games and Economic Behavior* 68(1): 339–52.

Trautmann, Stefan T. 2009. “A tractable model of process fairness under risk.” *Journal of Economic Psychology* 30(5): 803–13.

Weaver, R. Kent. 1986. “The Politics of Blame Avoidance.” *Journal of Public Policy* 6(4): 371–98.

Appendix A

Table A-1 reports the results of regression analyses to complement the non-parametric tests reported in the paper.

Column (1) shows an OLS regression of the punishment level on dummy variables of the dictator's decisions in the different treatments. Columns (2) and (3) report probit and logit regressions, respectively, of the punishment frequency on the same set of dummy variables. The omitted category in all regressions is the choice of the fair allocation in the baseline treatment.

The first three dummy variables measure the difference between the omitted category (fair) and the three other possible allocations in the baseline treatment. The unfair allocation is punished significantly more and more often than the fair allocation. (We also see that the dominated allocation is punished significantly more and more often, while the difference to the dominant allocation is small and partly insignificant.)

The next four dummy variables measure the difference between the baseline and the hidden information treatment when the dictator reveals the state. In all three regressions, none of the four coefficients is significant, which confirms our previous finding that the punishment pattern for a dictator who reveals is the same as the pattern in the baseline treatment.

Finally, the last four dummy variables measure the difference in punishment for a given allocation between a dictator who reveals and a dictator who remains ignorant in the hidden information treatment. All of the four comparisons in each of the three regression models show a significant difference of the previously reported sign, which again confirms the results of the non-parametric tests.

Table A-1: Regression Analyses

	(1) OLS	(2) Probit	(3) Logit
unfair (70-10)	19.17***	2.068***	3.710***
	(2.916)	(0.339)	(0.734)
dominant (70-50)	1.204	0.658**	1.354**
	(0.784)	(0.280)	(0.625)
dominated (50-10)	11.85***	1.786***	3.258***
	(2.064)	(0.331)	(0.722)
HI × unfair (70-10)	-3.472	-0.199	-0.318
	(3.926)	(0.238)	(0.382)
HI × dominant (70-50)	0.907	0.017	0.032
	(1.285)	(0.299)	(0.558)
HI × dominated (50-10)	-2.907	-0.341	-0.547
	(2.831)	(0.239)	(0.384)
HI × fair (50-50)	0.0278	0.141	0.314
	(0.605)	(0.420)	(0.937)
HI × ignorant × unfair (70-10)	-4.833***	-0.380***	-0.609***
	(1.257)	(0.133)	(0.214)
HI × ignorant × dominant (70-50)	5.333**	0.488**	0.860**
	(2.097)	(0.204)	(0.368)
HI × ignorant × dominated (50-10)	-3.500***	-0.232*	-0.381*
	(1.292)	(0.137)	(0.226)
HI × ignorant × fair (50-50)	3.833***	0.803***	1.558***
	(1.383)	(0.287)	(0.597)
Constant	0.556	-1.786***	-3.258***
	(0.472)	(0.319)	(0.724)
# observations	696	696	696
R-squared / Pseudo R-squared	0.156	0.142	0.142

Notes: The dependent variable in regression (1) is assigned punishment points by the third party. The dependent variable in regressions (2) and (3) is a dummy that equals 1 if the third party punishes. The omitted category in all regressions is the choice of the fair allocation in the baseline treatment. "HI" indicates the hidden information treatment. Robust standard errors clustering at the subject level are reported in parentheses. *** denotes significance at 1 percent, ** at 5 percent, and * at 10 percent.

Appendix B

B.1 Instructions for the dictators

We are pleased to welcome you to this economic study.

If you read the following instructions carefully, you can – depending on your decisions and/or those of the other participants – earn money in addition to the **15 Swiss francs** that you receive as an initial endowment for participating. It is thus very important that you read the instructions carefully. If you have any questions, please contact us.

During the study, speaking with the other participants in the study is forbidden. Violation of this rule will lead to exclusion from the study and all of the associated payments.

During the study, we will not speak of francs, but of points. Your entire income will thus first be calculated in points. The points you earn during the study will be converted to Swiss francs at the end of the study. The following conversion rate applies:

2 points = 1 Swiss franc.

At the end of today's study, you will receive the number of points earned during the study plus the initial endowment of 15 Swiss francs for appearing **in cash**.

We will explain the exact procedure of the study on the next pages. For the sake of simplicity, we will always use male forms; we obviously include female participants in any case.

The study

At the beginning of the study, you will be randomly paired with two other persons who are also participating in the study. You will neither learn during nor after the study of the identities of the two persons you are paired with. In the same way, the two other persons will never learn of your identity.

There are three types of participant in this study: participants A, B, and C. **You are a participant A.** The participants you are paired with are a participant B and a participant C.

This study lasts for **one** round. You thus only make your decisions one single time.

As participant A, you can decide how many points will be credited to you and to participant B.

In doing this, you can choose between **two possibilities**: You will always get 70 points for possibility 1 and always 50 points for possibility 2.

The points for participant B, however, depend on which **situation** you are in.

There are **two possible situations:**

- In situation 1, participant B receives 10 points for possibility 1 and 50 points for possibility 2.
- In situation 2, the payments are exactly opposite: participant B receives 50 points for possibility 1 and 10 points for possibility 2.

The table below gives you a summary of your selection possibilities and the points resulting in both situations. Please note that the points you receive are identical in both situations. Only the points for participant B are reversed.

	Situation 1		**Situation 2**	
	Points for participant A	Points for participant B	Points for participant A	Points for participant B
Possibility 1	70	10	70	50
Possibility 2	50	50	50	10

It will be randomly determined at the beginning of the study whether situation 1 or situation 2 is relevant. This information will not be automatically notified to you, however.

Your own payment is identical in situations 1 and 2. You can thus opt for possibility 1 or 2 without knowing or without later learning which situation is relevant.

However, you can **expose** the situation, and after doing this decide between possibility 1 and possibility 2.

If you **do not expose** the situation, you will not know which situation is relevant at the time of your decision between possibility 1 and 2, and you will **never learn of this**. You will also never learn which payment participant B received.

You thus make **two decisions** in this study:

1. You decide whether or not you want to expose which situation is relevant.
2. You decide between possibility 1 and possibility 2.

If you decide to expose the situation, we ask you to select a possibility **for both of the possible situations.** You can therefore – if you expose the situation – make your decision dependent on the actual situation. You will first learn which situation is actually relevant after you have made your decision for both possible situations. It is thus very important – if you expose the situation – that you consider your decisions in both situations very seriously.

Example 1: Participant A decides not to expose which situation is relevant. Afterwards, he opts for possibility 1. Participant A knows that he will receive 70 points, but he will never learn how many points participant B receives. He only knows that participant B will either receive 10 or 50 points, each with a probability of 50%.

Example 2: Participant A decides to expose which situation is relevant. He decides that he will select possibility 2 in situation 1 and possibility 1 in situation 2. He then learns that situation 2 applies. His decision in situation 2 will thus be implemented. Participant A knows that he will earn 70 points and that participant B will receive 50 points.

Participant B cannot make any decisions on his own in this study.

Participant C has an initial endowment of **50 points**. He observes

(i) whether you expose the situation or not;

(ii) which situation is relevant (even if you do not expose the situation);

(iii) and which possibility you select in the relevant situation.

Participant C has the opportunity to deduct points from you. Participant C can determine the point deduction in units of 5. He can thus deduct 0, 5, 10, 15, etc. points.

A deduction of a unit of 5 points costs participant C one point. If, for example, participant C deducts 15 points from you, his own payment will be reduced by 3 points.

Participant C can only deduct a maximum number points that will reduce your payment to 10 points. Participant C can also decide not to deduct any points from you.

Example 3: Participant C observes (i) that you expose the situation; (ii) that situation 2 is relevant; and (iii) that you selected possibility 1. He decides to deduct 10 points from you. The following payments thus result:

Points for participant A	Points for participant B	Points for participant C
70 – 10 = 60	50	50 – 2 = 48

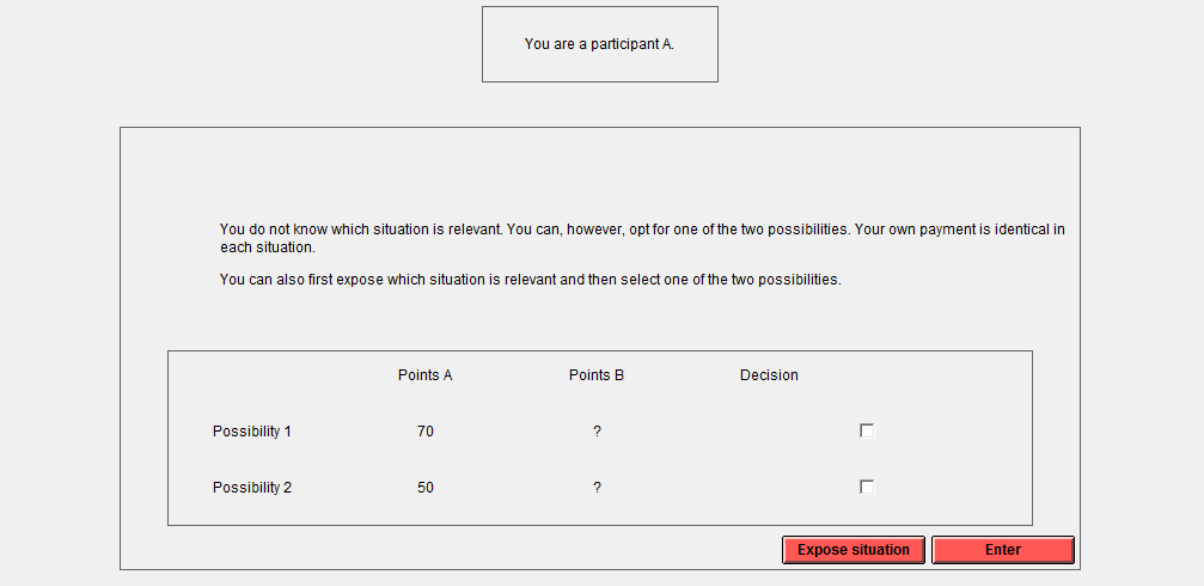
Example 4: Participant C observes (i) that you do not expose the situation. (ii) Unlike you, he observes that situation 1 is relevant, and (iii) that you selected possibility 1. He decides not to lose any points and thus not to deduct points from you. The following payments thus result:

Points for participant A	Points for participant B	Points for participant C
70	10	50

Please take enough time to become familiar with the decision possibilities. It is very important that you understand the decision possibilities exactly, as you can only make the two decisions "expose or not expose" and "possibility 1 or 2" once, and that the study concludes after this.

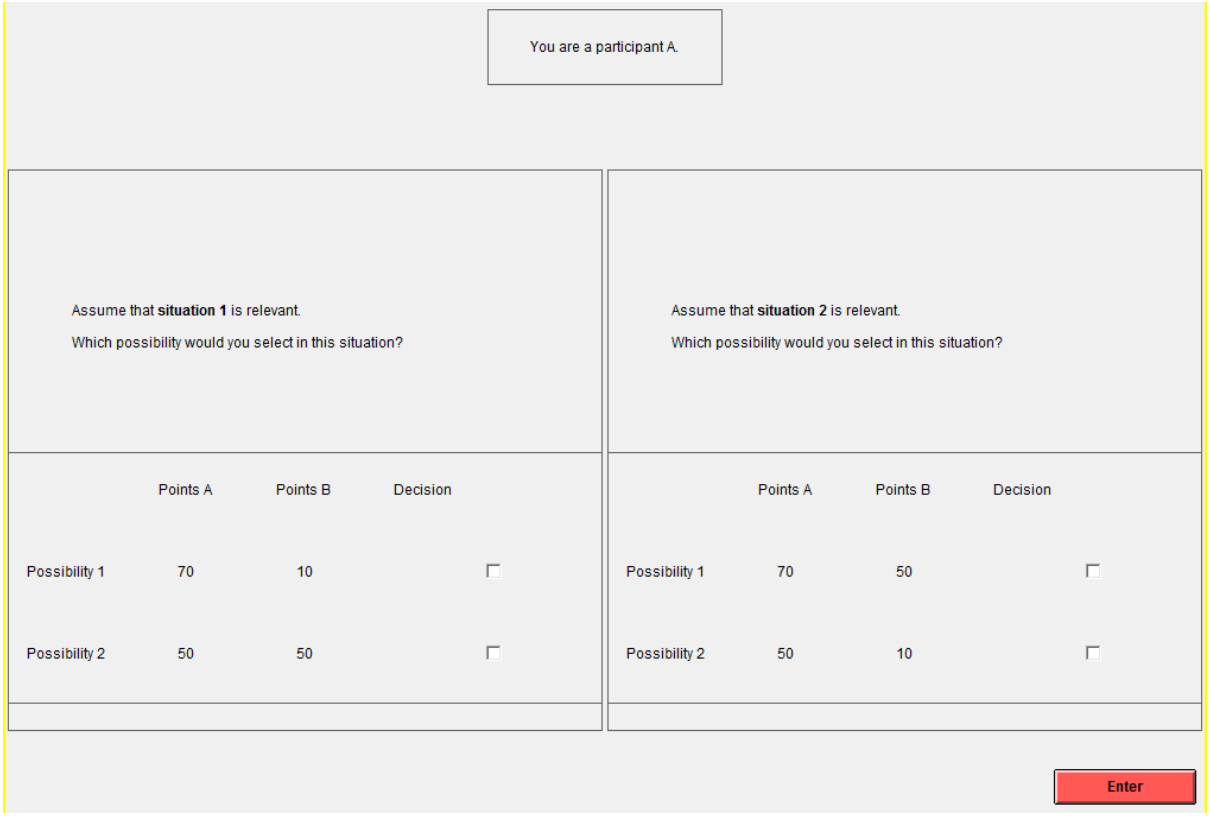
Procedure on the computer:

You make your decision as participant A on the screen shown below. You see your payment in case of possibility 1 and 2. However, you do not see the payment that results for participant B.



If you want to select possibility 1 or 2 without exposing the situation in advance, then please click on the corresponding box for possibility 1 or 2. In order to confirm your decision, you must also click on the enter button at the bottom right-hand side. You can change your decision until you click this button.

If you want to first expose the situation, you must click on the button "Expose situation". If you decide to expose the situation, the screen below appears. You see the both possibilities in situation 1 on the left side of the screen and both possibilities in situation 2 on the right side of the screen.



You make your decision between the two possibilities by clicking the corresponding box in each of the situations. You will not learn which of the two situations is relevant for the payment until you have made your decisions for both situations.

Once you have made your decisions for both situations, click on the enter button at the lower right hand side. You can change your decision until you click this button.

At the end of the study, the number of points you earn will be converted to Swiss francs and, together with the initial endowment, paid out in cash to you.

Do you have any questions? If yes, please raise your hand. We will then come to you at your workplace. If everything is clear, we kindly ask you to answer the control questions on the following pages.

Control questions

Please answer the following control questions. They only have the purpose of making you more familiar with the study. Your answers do not influence your payment at the end of the study.

1. Participant A does not expose the situation and selects possibility 1. How many points does participant A receive, and how many does participant B receive with a probability of 50% each?

2. Participant A does not expose the situation and selects possibility 2. How many points does participant A receive, and how many does participant B receive with a probability of 50% each?
3. Participant A exposes the situation. Situation 1 is relevant, and participant A selected possibility 2 in situation 1. How many points to participant A and participant B receive in this case?
4. Participant A exposes the situation. Situation 2 is relevant, and participant A selected possibility 1 in situation 2. How many points to participant A and participant B receive in this case?
5. Participant A does not expose the situation and selects possibility 2. Participant C learns that situation 2 is relevant. What is the maximum number of points that participant C can deduct from participant A?
6. Participant A exposes the situation. Situation 1 is relevant, and participant A selected possibility 1 in situation 1. What is the maximum number of points that participant C can deduct from participant A?
7. Participant A does not expose the situation and selects possibility 2. Participant C learns that situation 2 is relevant. In this case, participant C does not deduct any points from participant A. Please determine the final payments for participants A, B, and C.

Points for participant A	Points for participant B	Points for participant C

8. Participant A exposes the situation. Situation 1 is relevant, and participant A selected possibility 1 in situation 1. Participant C deducts 30 points from participant A. Please determine the final payments for participants A, B, and C.

Points for participant A	Points for participant B	Points for participant C

9. Participant A exposes the situation. Situation 2 is relevant, and participant A selected possibility 1 in situation 2. Participant C deducts 60 points from participant A. Please determine the final payments for participants A, B, and C.

Points for participant A	Points for participant B	Points for participant C

Please raise your hand when you have solved all the questions. We will then come to your workplace and check your answers.

B.2 Instructions for the receivers / third parties

We are pleased to welcome you to this economic study.

If you read the following instructions carefully, you can – depending on your decisions and/or those of the other participants – earn money in addition to the **15 Swiss francs** that you receive as an initial endowment for participating. It is thus very important that you read the instructions carefully. If you have any questions, please contact us.

During the study, speaking with the other participants in the study is forbidden. Violation of this rule will lead to exclusion from the study and all of the associated payments.

During the study, we will not speak of francs, but of points. Your entire income will thus first be calculated in points. The points you earn during the study will be converted to Swiss francs at the end of the study. The following conversion rate applies:

2 points = 1 Swiss franc.

At the end of today's study, you will receive the number of points earned during the study plus the initial endowment of 15 Swiss francs for appearing **in cash**.

We will explain the exact procedure of the study on the next pages. For the sake of simplicity, we will always use male forms; we obviously include female participants in any case.

The study

At the beginning of the study, you will be randomly paired with two other persons who are also participating in the study. You will neither learn during nor after the study of the identities of the two persons you are paired with. In the same way, the two other persons will never learn of your identity.

There are three types of participant in this study: participants A, B, and C. **You are either a participant B <u>or</u> a participant C.** You will not learn until the end of the study whether you are a participant B or a participant C.

Each group consists of one participant A, B, and C each. If you are a participant B, you will thus be paired with a participant A and a participant C. If, however, you are a participant C, you will thus be paired with a participant A and a participant B.

This study lasts for **one** round. You thus only make your decisions one single time.

Participant A can decide how many points will be credited to himself and to participant B.

In doing this, Participant A chooses between **two possibilities**: He will always get 70 points for possibility 1 and always 50 points for possibility 2.

The points for participant B, however, depend on which **situation** Participant A is in.

There are **two possible situations:**

- In situation 1, participant B receives 10 points for possibility 1 and 50 points for possibility 2.
- In situation 2, the payments are exactly opposite: participant B receives 50 points for possibility 1 and 10 points for possibility 2.

The table below gives you a summary of participant A's selection possibilities and the points resulting in both situations. Please note that the points participant A receives are identical in both situations. Only the points for participant B are reversed.

	Situation 1		**Situation 2**	
	Points for participant A	Points for participant B	Points for participant A	Points for participant B
Possibility 1	70	10	70	50
Possibility 2	50	50	50	10

It will be randomly determined at the beginning of the study whether situation 1 or situation 2 is relevant. This information will not be automatically notified to participant A, however.

Participant A's payment is identical in situations 1 and 2. He can thus opt for possibility 1 or 2 without knowing or without later learning which situation is relevant.

However, participant A can **expose** the situation, and after doing this decide between possibility 1 and possibility 2.

If participant A **does not expose** the situation, he will not know which situation is relevant at the time of his decision between possibility 1 and 2, and he will **never learn of this**. Participant A will also never learn which payment participant B received.

Participant A thus makes **two decisions** in this study:

1. He decides whether or not he wants to expose which situation is relevant.
2. He decides between possibility 1 and possibility 2.

Example 1: Participant A decides not to expose which situation is relevant. Afterwards, he opts for possibility 1. Participant A knows that he will receive 70 points, but he will never learn how many points participant B receives. He only knows that participant B will either receive 10 or 50 points, each with a probability of 50%.

Example 2: Participant A decides to expose which situation is relevant. Situation 2 is relevant, and participant A selects possibility 1. Participant A knows that he will earn 70 points and that participant B will receive 50 points.

Participant B cannot make any decisions on his own in this study.

Participant C has an initial endowment of **50 points**.

Participant C has the opportunity to deduct points from participant A. Participant C can determine the point deduction in units of 5. He can thus deduct 0, 5, 10, 15, etc. points.

A deduction of a unit of 5 points costs participant C one point. If, for example, participant C deducts 15 points from participant A, his own payment will be reduced by 3 points.

Participant C can only deduct a maximum number points from participant A that will reduce the latter's payment to 10 points. Participant C can also decide not to deduct any points from participant A.

Participant A makes his decisions without knowing whether participant C will deduct points from him or not. Participant A only learns of this at the end of the study.

As you do not yet know whether you are a participant B or a participant C, we ask you to indicate which decision you would make as participant C. If you are actually a participant C, your decision will be implemented.

Before you observe

(iv) whether participant A exposes the situation or not;

(v) which situation is relevant (even if participant A does not expose the situation);

(vi) and which possibility participant A selects in the relevant situation;

we ask you to indicate how many points you would deduct from participant A **in each possible case** in case you are a participant C.

There are eight possible cases:

- Four cases result when participant A exposes the situation. Situation 1 or situation 2 can then be relevant. In each situation, participant A can decide between possibility 1 and possibility 2

  Participant A exposes the situation:

Situation 1 is relevant	
Case: Participant A selects possibility 1	**How many points would you deduct?**
Case: Participant A selects possibility 2	**How many points would you deduct?**

Situation 2 is relevant	
Case: Participant A selects possibility 1	**How many points would you deduct?**
Case: Participant A selects possibility 2	**How many points would you deduct?**

- Four cases result when participant A does **not** expose the situation. Participant A decides between possibility 1 and 2 **without knowing which situation is relevant.** As you – unlike participant A – know which situation is relevant, you can make your point deduction dependent on the relevant situation.

  Participant A does **not** expose the situation and thus does not know which situation is relevant:

Situation 1 is relevant	
Case: Participant A selects possibility 1 <u>without knowing how this affects participant B</u>	**How many points would you deduct?**
Case: Participant A selects possibility 2 <u>without knowing how this affects participant B</u>	**How many points would you deduct?**

Situation 2 is relevant	
Case: Participant A selects possibility 1 <u>without knowing how this affects participant B</u>	**How many points would you deduct?**
Case: Participant A selects possibility 2 <u>without knowing how this affects participant B</u>	**How many points would you deduct?**

You will first learn whether you are a participant B or a participant C after you have made your decision in all eight cases. You will also then learn (i) whether participant A exposes the situation, (ii) which situation is relevant, and (iii) which possibility participant A selects. **<u>Each of your eight decisions can thus determine your payment.</u>**

Example 3: You learn that you are a participant C and you observe (i) that participant A exposes the situation; (ii) that situation 2 is relevant; and (iii) that he selected possibility 1. In this case, you decided to deduct 10 points from participant A. The following payments thus result:

Points for participant A	Points for participant B	Points for participant C
70 – 10 = 60	50	50 – 2 = 48

Example 4: You learn that you are a participant B, and you observe (i) that participant A does not expose the situation. (ii) Unlike participant A, you observe that situation 1 is relevant, and (iii) that participant A selected possibility 1. The participant C assigned to you decides not to lose any points and thus not to deduct points from participant A. The following payments thus result:

Points for participant A	Points for participant B	Points for participant C
70	10	50

Please take enough time to become familiar with the decision possibilities. It is very important that you understand the decision possibilities exactly, as you can only make a decision about point deductions in the eight possible cases once, and that the study concludes after this.

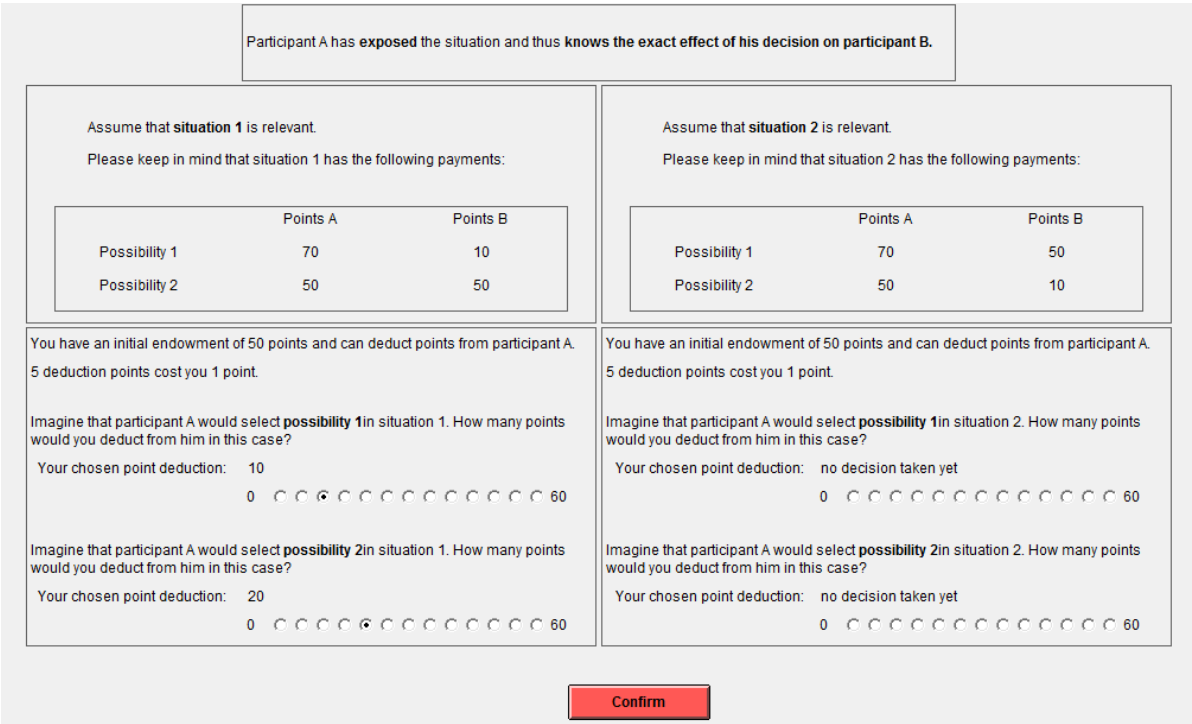
Procedure on the computer:

You enter your decisions on the screen below. In the upper part of the screen, you see whether the shown cases are those where participant A has exposed the situation or not. Here you see the screen for the case where participant A has exposed the situation.

You enter your reaction in situation 1 on the left side of the screen, and you enter your reaction to situation 2 on the right side of the screen. Please enter in both situations how you would react if you were participant C to possible decisions by participant A.

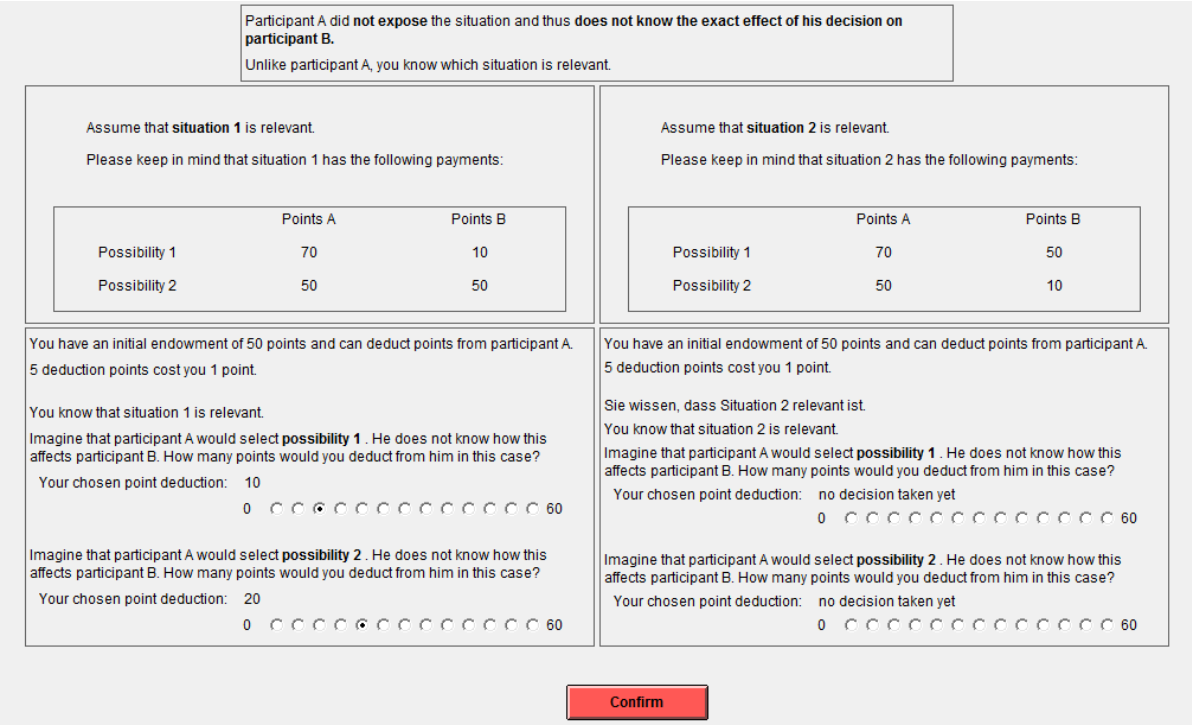
You enter the desired point deduction by clicking on the corresponding field. The field at the very left corresponds to a point deduction of 0, and that at the very right to a point deduction of 60. Any unit of five between these numbers is also possible. The point deduction will always appear to the right, next to “Your chosen point deduction”.

The screen below shows a **randomly selected example**; participant C deducts 10 points if participant A selects possibility 1 in situation 1 and 20 points if he selects possibility 2. Participant C has not yet made a decision for situation 2.



After you have selected your deduction points for all four possibilities, please click on the "confirm" button. You can change your decision until you click this button.

The screen for the case where participant A does not expose the situation is very similar. Here you see an example:



The order in which the screens appear is random. **Please take exact note therefore of the case for which you make your decisions!**

At the end of the study, the number of points you earn will be converted to Swiss francs and, together with the initial endowment, paid out in cash to you.

Do you have any questions? If yes, please raise your hand. We will then come to you at your workplace. If everything is clear, we kindly ask you to answer the control questions on the following pages.

Control questions

Please answer the following control questions. They only have the purpose of making you more familiar with the study. Your answers do not influence your payment at the end of the study.

1. Participant A does not expose the situation and selects possibility 1. How many points does participant A receive, and how many does participant B receive with a probability of 50% each?
2. Participant A does not expose the situation and selects possibility 2. How many points does participant A receive, and how many does participant B receive with a probability of 50% each?
3. Participant A exposes the situation. Situation 1 is relevant, and participant A selected possibility 2 in situation 1. How many points to participant A and participant B receive in this case?
4. Participant A exposes the situation. Situation 2 is relevant, and participant A selected possibility 1 in situation 2. How many points to participant A and participant B receive in this case?
5. Participant A does not expose the situation and selects possibility 2. Participant C learns that situation 2 is relevant. What is the maximum number of points that participant C can deduct from participant A?
6. Participant A exposes the situation. Situation 1 is relevant, and he selected possibility 1 in situation 1. What is the maximum number of points that participant C can deduct from participant A?
7. Participant A does not expose the situation and selects possibility 2. Participant C learns that situation 2 is relevant. In this case, participant C does not deduct any points from participant A. Please determine the final payments for participants A, B, and C.

Points for participant A	Points for participant B	Points for participant C

8. Participant A exposes the situation and learns that situation 1 is relevant. He selects possibility 1. Participant C deducts 30 points from participant A in this case. Please determine the final payments for participants A, B, and C.

Points for participant A	Points for participant B	Points for participant C

9. Participant A exposes the situation and learns that situation 2 is relevant. He selects possibility 1. Participant C deducts 60 points from participant A in this case. Please determine the final payments for participants A, B, and C

Points for participant A	Points for participant B	Points for participant C

Please raise your hand when you have solved all the questions. We will then come to your workplace and check your answers.