Three Economic Experiments on Norm Enforcement, Honesty, and Strategic Gaze

Dissertation

zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften
(Dr. rer. pol.)

vorgelegt von:

Konstantin von Hesler

Universität Konstanz

Sektion Politik-Recht-Wirtschaft
Fachbereich Wirtschaftswissenschaften

Konstanz, 2017

Tag der mündlichen Prüfung: 14.12.2017

1. Referent: Prof. Dr. Urs Fischbacher

2. Referent: Prof. Dr. Peter Gollwitzer

3. Referent: Prof. Dr. Gerald Eisenkopf

Für Johanna

Danksagung

Auf meinem zehnjährigen Weg durch Konstanz wurde ich von unzähligen Menschen begleitet und unterstützt. Dafür möchte ich mich an dieser Stelle bedanken.

An erster Stelle gilt mein Dank meinem Doktorvater Urs Fischbacher. Unser gemeinsamer Weg fing schon in meinem ersten Studienjahr in Konstanz an, da Urs damals (ebenfalls in seinem ersten Jahr in Konstanz) die Mikro I Vorlesung von Friedrich Breyer übernommen hatte, der sich im Forschungssemester befand. Urs hat in dieser Vorlesung zunächst mein Interesse für die Mikroökonomik geweckt und ich habe daraufhin angefangen, seine Veröffentlichungen zu lesen. Dadurch ist schon früh meine bis heute lodernde Begeisterung für die experimentelle Wirtschaftsforschung geweckt worden. Seit der Mikro I hat Urs mich akademisch begleitet und betreut—so habe ich nicht nur meine Doktorarbeit unter seiner Betreuung geschrieben, sondern auch meine Bachelor- sowie meine Masterarbeit. Urs hat mich über all die Jahre stets großzügig unterstützt und immer wieder meinen Enthusiasmus neu entfacht. Für all das: vielen, vielen Dank!!

Ich möchte auch meinem zweiten Betreuer Gerald Eisenkopf danken. Auch er hat mich inzwischen über viele Jahre begleitet und stand mir nicht nur mit akademischem Rat zur Seite, sondern wusste auch immer besonders gut, jeden Moment mit seiner Fröhlichkeit zu verschönern. Ich werde nie dein schallendes Lachen im Nachbarbüro vergessen! Vielen Dank, Gerald!

Auch meinem dritten Betreuer Peter Gollwitzer möchte ich besonders danken. Er war es, der mich schon in meinem ersten Studienjahr besonders ermutigt hat, mich in die Forschungskolloquien zu setzen und mich mit Urs in Verbindung zu setzen. In meiner Zeit in New York habt ihr mich so familiär aufgefangen und selbstlos gestützt—ohne euch hätte ich es nicht geschafft. Dafür unendlichen Dank!

Ich bedanke mich außerdem von ganzem Herzen bei all meinen Freunden, die mich in den letzten Jahren begleitet haben. Ein Ehrenplatz gebührt Jan Hausfeld, mit dem ich wirklich jede unserer Ideen—seien sie noch so blöd—völlig frei zu Ende denken kann. Die Ergebnisse zweier unserer besseren Ideen befinden sich in Kapitel 2 und 3. Jedes Gespräch mit dir ist eine Bereicherung. Es gibt nur wenige, mit denen ich so gut reden, feiern, dumme Videos gucken oder einfach nur chillen und nichts tun kann wie mit dir. 1000 Dank!

Im gleichen Atemzug bedanke ich mich bei Thomas Hattenbach und Kati. Danke für unzählige Abendessen, Fail-Video-Compilations sowie für stundenlange Gespräche über Gott und die Welt! Für endlose Nächte, Skihütten, Wakeboardtouren, Gunf, den gemeinsamen Weg und allgemeine Exzesse bedanke ich mich auch bei Domink (sic!), Carl, Chris, Schlecki, Baiba, Klāvs, Nathalie, Franzi, David, Fabi, Micha, Mika, Philipp und Silvana. Mein Dank gilt auch meinen engsten Freunden aus früheren Episoden: Danke an Simon Hess, Felix Steiner und Fabio Bröcker! Mit euch ist alles besser! Vielen Dank an meine liebste Cousine Nandi. Auch bei Sina möchte ich mich an dieser Stelle ganz besonders bedanken, die jede meiner Endspurt-Marotten ausdauernd ertragen und mit viel Nachsicht über sie hinweggesehen hat.

Ich bedanke mich außerdem bei der stetig gewachsenen Gruppe am TWI für die Freundschaft und den intensiven Austausch: Vielen Dank an Dominik Bauer, David Dohmen, Fabian Dvorak, David Grammling, Thomas Hattenbach, Jan Hausfeld, Moritz Janas, Baiba Renerte, Katrin Schmelz, Irenaeus Wolff sowie auch an unsere ehemaligen Kollegen Fadong Chen, Kate Bendrick, Lisa Bruttel, Simeon Schudy und Pascal Sulser. Auch Barbara Monstein, Madeleine Hafner, Jutta Obenland und Justine Overall gebührt besonderer Dank! Auch bei Susanne Goldlücke möchte ich mich ausdrücklich für unsere Zusammenarbeit bedanken.

Vielen Dank auch an das TWI selbst, die Deutsche Forschungsgemeinschaft, die Graduate School of Decision Sciences sowie den Young Scholar Fund, die alle maßgeblich an der Finanzierung dieser Dissertation beteiligt waren.

Abschließend möchte ich mich bei meiner engsten Familie bedanken. Es lässt sich nicht in Worte fassen, wie sehr ihr mich stützt. 1000 Dank an Mami & Papi, Carly, Nik, Raschi und Johanna.

Contents

Summary

In this dissertation, I present three economic experiments in the chronological order in which I started working on the respective projects. The first is a collaboration with my thesis adviser, Urs Fischbacher, in which we investigated when and why unaffected third parties are willing to punish and reward decision makers after observing their behavior toward others. The second study is joint work with Jan Hausfeld. We explored the effects of alcohol intoxication on honesty in a lab-in-the-field experiment. In my third project, conducted jointly with Susanne Goldlücke and Jan Hausfeld, we studied if and how people actively control their eye movements in order to gain a strategic advantage. I will give a more detailed summary of each individual paper in the following three paragraphs.

The first paper is entitled “Processes and Preferences Underlying Third Party Rewards and Punishments”. We used an experiment to investigate how third parties decide on punishments and rewards in a series of mini dictator games. In each of these games, a dictator chooses between two possible allocations of ten points between himself and a recipient. Without knowing the dictator’s decision, a mechanism *designer* decides whether punishment or reward will be made available to the mechanism *implementer* who does observe the dictator’s decision and then decides whether or not to use the available mechanism. We find that implementers are less likely to punish and more likely to reward the higher the dictator’s transfer to the recipient and the better her revealed intention. Similarly, mechanism designers are more likely to provide the reward mechanism if the dictator’s expected decision is relatively more worthy of reward than punishment. These findings are also reflected in the mechanism designers’ decision times which reveal that their decisions are faster when the dictator’s expected behavior is either clearly worthy of punishment (but not reward) or of reward (but not punishment).

In the second paper, “In Vino Veritas”, we studied the effect of social drinking on honesty in a lab-in-the-field experiment. We recruited bar-goers before they entered a bar and—depending on treatment—they either participated in the experiment right away (sober condition) or when they came back out from the bar (intoxicated condition). In both conditions, participants’ degree of intoxication was measured with a breathalyzer before and after the bar visit. This design allows us to disentangle the effect of acute alcohol intoxication from that of general drinking habits by using the randomly imposed treatment condition as an instrumental variable for the degree of intoxication during the experiment. At the same time, the two breathalyzer measurements are used as a proxy for general drinking habits. In our game to measure honesty, participants privately draw a tile from a pouch and then report the outcome of their draw to determine their earnings. Since only they know the true outcome, they can lie in order to get a higher payoff. As in Fischbacher & Föllmi-Heusi (2013), the amount of dishonesty is determined only statistically by comparing the observed amount of reported high outcomes with the one expected under full honesty. We find that people become more honest the more they had to drink, while at the same time people who drink more in general are also less honest in general.

The third paper, “Strategic Gaze: An Interactive Eye-Tracking Study”, explores the strategic use of gaze by analyzing eye movements and their consequences in four simple games. The game can either be a competitive (hide & seek) game in which players want to be unpredictable, or a game of common interest in which players want to be predictable. Gaze is either transmitted in real time to another subject, or it is not transmitted and therefore non-strategic. We find that subjects are able to interpret previously recorded non-strategic gaze, obtaining substantially higher payoffs than subjects who did not see gaze patterns. If gaze is transmitted in real time, eye-movements become even more informative in the common interest games and players predominantly succeed to coordinate on efficient outcomes. Also in the hide & seek game, people are able to correctly predict choice from non-strategic gaze while this is no longer the case when gaze is strategic. With live gaze transmission, most eye-tracked players successfully jam their signals either by quickly looking back and forth between the available options or by simply not looking at all. Some of them also try to actively trick the other player into holding a wrong belief about the intended choice by emphasizing one of the options with their gaze. However, they are not more successful on average than those who use the signal jamming strategy.

Zusammenfassung

In dieser Dissertation präsentiere ich drei ökonomische Experimente in der Reihenfolge, in der ich mit der Arbeit an den entsprechenden Projekten begonnen habe. Das erste ist eine Kollaboration mit meinem Doktorvater Urs Fischbacher, in der wir untersucht haben, wann und warum unbeteiligte Dritte dazu bereit sind, jemanden für seine Entscheidungen zu belohnen oder zu bestrafen, nachdem sie dessen Verhalten gegenüber anderen beobachtet haben. Die zweite Studie entstand in Zusammenarbeit mit Jan Hausfeld, in der wir mit einem *lab-in-the-field* Experiment den Einfluss von Alkoholkonsum auf Ehrlichkeit untersucht haben.[1] In meinem dritten Projekt, das ich gemeinsam mit Susanne Goldlücke und Jan Hausfeld durchgeführt habe, haben wir untersucht, ob und wie Menschen ihre Augenbewegungen aktiv steuern, um sich einen strategischen Vorteil zu verschaffen. In den folgenden drei Absätzen werde ich jeden der drei Artikel detaillierter zusammenfassen.

Der erste Artikel trägt den Titel „Processes and Preferences Underlying Third Party Rewards and Punishments" („Zugrundeliegende Prozesse und Präferenzen von Belohnungen und Bestrafungen durch unbeteiligte Dritte"). Mit Hilfe eines Experiments haben wir untersucht, wie unbeteiligte Dritte über Belohnungen und Bestrafungen in einer Reihe von Mini-Diktator-Spielen entscheiden. In jedem dieser Spiele wählt ein Diktator eine von zwei möglichen Aufteilungen von zehn Punkten zwischen sich selbst und einem Empfänger. Ohne die Entscheidung des Diktators zu kennen entscheidet ein Mechanismus-*Designer*, ob dem Mechanismus-*Umsetzer* eine Möglichkeit zur Belohnung oder eine Möglichkeit zur Bestrafung des Diktators zur Verfügung gestellt wird. Der Mechanismus-Umsetzer beobachtet anschließend die Entscheidung des Diktators und entscheidet dann, ob der verfügbare Mechanismus eingesetzt werden soll oder

[1] *Lab-in-the-field* Experimente sind eine Mischung aus Labor- und Feldexperiment. Sie werden zwar im Feld und mit den dort vorhandenen Teilnehmern durchgeführt, abgesehen davon herrschen jedoch ähnliche Verhältnisse wie in einem Labor.

nicht. Unsere Ergebnisse zeigen: Je höher der Transfer vom Diktator zum Empfänger und je besser seine durch seine Wahl offenbarte Intention, desto unwahrscheinlicher wird eine Bestrafung und desto wahrscheinlicher wird eine Belohnung durch den Mechanismus-Umsetzer. Ähnlicherweise gilt auch, dass der Mechanismus-Designer mit größerer Wahrscheinlichkeit den Belohnungsmechanismus bereitstellen wird, je eher die erwartete Entscheidung des Diktators eine Belohnung statt einer Bestrafung verdient hat. Diese Ergebnisse spiegeln sich auch in den Reaktionszeiten der Mechanismus-Designer wider: Ihre Entscheidungen sind schneller, wenn das erwartete Verhalten der Diktatoren entweder eindeutig eine Belohnung (aber keine Bestrafung) oder eine Bestrafung (aber keine Belohnung) verdient.

Im zweiten Artikel, „In Vino Veritas“, haben wir den Einfluss von Alkoholkonsum im sozialen Umfeld auf Ehrlichkeit in einem *lab-in-the-field* Experiment untersucht. Hierfür haben wir Barbesucher rekrutiert, direkt bevor sie in eine Bar gehen wollten. Abhängig von der Experimental-Bedingung haben sie entweder direkt nach der Rekrutierung am Experiment teilgenommen (nüchterne Bedingung) oder nach dem sie die Bar wieder verlassen haben (alkoholisierte Bedingung). In beiden Bedingungen wurde der Grad der Alkoholisierung jedes Teilnehmers je zwei mal mit einem Atemalkoholmessgerät gemessen: einmal vor und einmal nach dem Barbesuch. Mit Hilfe dieses Designs können wir den Einfluss akuter Alkoholisierung von dem der allgemeinen Trinkgewohnheiten unterscheiden, indem wir die zufällig zugeteilte Experimental-Bedingung als Instrumentvariable für den Alkoholisierungsgrad während des Experiments benutzen. Gleichzeitig können die zwei Atemalkoholmessungen als Proxy-Variable für die allgemeinen Trinkgewohnheiten genutzt werden. Wir messen Ehrlichkeit mit einem Spiel in dem die Teilnehmer nur für sich selbst sichtbar einen Spielstein aus einem Beutel ziehen, anschließend das Ergebnis eingeben und damit ihr Einkommen bestimmen. Da nur die Teilnehmer selbst das wahre Ergebnis kennen, können sie dabei lügen, um ein höheres Einkommen zu erzielen. Wie bei Fischbacher & Föllmi-Heusi (2013) wird der Grad an Unehrlichkeit lediglich statistisch bestimmt, indem die beobachtete Anzahl der von den Teilnehmern angegebenen hohen Einkommen mit der verglichen wird, die bei vollständiger Ehrlichkeit zu erwarten gewesen wäre. Wir finden einerseits, dass Leute ehrlicher werden, je mehr sie getrunken haben, und andererseits, dass Leute die im Allgemeinen mehr trinken auch allgemein unehrlicher sind.

Der dritte Artikel, „Strategic Gaze: An Interactive Eye-Tracking Study" („Strategische Blicke: eine interaktive Eye-Tracking Studie") untersucht den strategischen Einsatz von Blicken, indem Augenbewegungen und deren Konsequenzen in vier einfachen Spielen untersucht werden. Das Spiel kann entweder ein kompetitives Versteckspiel sein, in dem die Spieler möglichst unvorhersehbar sein möchten oder ein Spiel mit geteilten Interessen, in dem die Spieler möglichst vorhersehbar sein möchten. Das Blickverhalten wird entweder in Echtzeit von einem der Spieler zu seinem Partner übertragen oder es wird nicht übertragen und ist deswegen nicht strategisch. Wir finden, dass Teilnehmer in der Lage sind, zuvor aufgenommenes, nicht strategisches Blickverhalten zu interpretieren, sodass sie substantiell höhere Einkommen erzielen als andere, die nicht sehen konnten, wo ihr Partner hinsieht. Wenn die Blicke in Echtzeit übertragen werden, werden die Augenbewegungen in Spielen mit geteilten Interessen sogar noch informativer und die Spieler schaffen es mehrheitlich, sich auf effiziente Lösungen zu verständigen. Auch das nicht strategische Blickverhalten aus dem Versteckspiel wird von den Teilnehmern verstanden. Bei strategischem Blickverhalten ist dies jedoch nicht mehr der Fall. Bei Liveübertragung der Blickdaten blockieren die meisten Teilnehmer mit Eye-Trackern ihre Signale, indem sie entweder schnell zwischen den verfügbaren Optionen hin und her sehen, oder indem sie einfach entscheiden, ohne überhaupt etwas anzusehen. Manche versuchen auch aktiv, den anderen Spieler in die Irre zu führen, indem sie eine der Optionen durch ihre Blicke besonders hervorheben, jedoch sind diese Spieler im Durchschnitt nicht erfolgreicher als die anderen.

Processes and Preferences Underlying Third Party Rewards and Punishments

By Urs Fischbacher & Konstantin Hesler

We use an experiment to investigate how unaffected third parties decide on punishments and rewards in a series of mini dictator games. For each game, a mechanism designer decides whether punishment or reward will be available to the mechanism implementer who observes the dictator's decision and then decides whether or not to use the available mechanism. Implementers are less likely to punish and more likely to reward the higher the dictator's transfer to the recipient and the better her revealed intention. Similarly, mechanism designers are more likely to provide the reward mechanism if the dictator's expected decision is relatively more worthy of reward than punishment. These findings are also reflected in the mechanism designer's decision times which reveal that their decisions are faster when the dictator's expected behavior is either clearly worthy of punishment or of reward. Decision times of the mechanism implementers are less conclusive.

JEL Codes: *C91, D03, D63*

Keywords: *Punishment, reward, third party, dictator game, equality, intention*

I. Introduction

When are unaffected third parties willing to punish or reward people for their behavior toward others? We try to answer this question with an experiment. More specifically, we let our participants play a series of ten constant-sum mini dictator games with two third parties: One mechanism designer and one mechanism implementer. Without knowing the dictator's choice, the designer decides for each game whether punishment or reward will be available as a mechanism to sanction the dictator. The implementer observes both the decision of the designer and that of the dictator and then decides whether she wants to apply the available mechanism or not. Applying a mechanism comes at a small cost to the implementer but is costless to the designer.

Our findings are intuitive: The higher the transfer from the dictator to the recipient, the more likely will an implementer reward and the less likely will she punish. Similarly, the better the dictator's revealed intention, the more likely will she be rewarded and the less likely will she be punished. We further test predictions regarding the implementers' decision times based on the evidence accumulation model (Ratcliff, 1978). When implementers can punish selfish dictators, their decision times go as predicted, i.e., they are slowest around an intermediate revealed intention and an intermediate transfer and they become faster, the further away from these thresholds. Our results are puzzling for the case of rewards however: We find exactly the opposite of the predicted effect of the dictator's transfer on the implementers' reward decision times, i.e., they are *fastest* for an intermediate transfer and become slower for more extreme values. Further, there is no significant effect of the revealed intention on the reward decision times. These results indicate that, even though humans seem to draw on the same general motives for reward and punishments, the two decisions seem to be processed in a fundamentally different way.

Similarly, we investigate how mechanism designers make their decisions depending on the dictators' potential worthiness of punishment or reward and the potential efficiency effects of the two mechanisms (based on the expected costs to the mechanism implementers). We use three different measures for the dictators' potential worthiness of punishment/reward—one based on the designers' belief about the dictators' behavior to capture the potential intentions, one based on the dictator's expected transfer, and one based on the punishment and reward decisions of the mechanism implementers. Our data show that the dictators' potential worthiness of punishment/reward clearly dominates any efficiency concerns. In fact, our regressions show that, because of the negative correlation between expected efficiency effects and the dictator's potential worthiness of punishment/reward, the designers even significantly decrease expected efficiency. This interpretation finds additional support in the analysis of the designers' decision times which shows no significant effect of the efficiency effects but an effect as predicted by the evidence accumulation model for the potential worthiness of punishment/reward.

Unlike this study, most previous research on norm-enforcement has focused on peer punishments and their efficacy to promote cooperation (see, e.g., Fehr & Gächter, 2000; Fehr & Gächter, 2002; Gächter, Renner & Sefton, 2008 and the literature reviewed by Milinski & Rockenbach, 2012). It has been argued however, that third party interventions are a crucial societal feature to enable norms to evolve and be sustained (Bendor & Swistak, 2001; Buckholtz & Marois, 2012), especially so in larger societies (Marlowe et al., 2008; Henrich et al., 2010). The question when and why third parties decide to punish and reward is thus highly relevant since these sanctions can only have their positive impact on society if they are in fact executed. In the U.S. alone, each year millions of cases are brought before a judge who has to decide whether there has been a norm transgression worthy of punishment or not. Likewise, there are countless awards and honors presented for various achievements every year.

Experimental economists have previously established that third parties indeed punish and reward even if it is costly (Kahneman, Knetsch & Thaler, 1986; Turillo et al., 2002; Fehr & Fischbacher, 2004; Charness, Cobo-Reyes & Jiménez, 2008; Almenberg et al., 2011; de Kwaadsteniet, Rijkhoff & van Dijk, 2013; Bartling, Engl & Weber, 2014; and Nikiforakis & Mitchell, 2014).[1] With

[1] A further strand of related literature is concerned with cultural and societal determinants of third party punishments. Henrich et al. (2006) found that a demand for costly third party punishments exists in a wide variety of subject pools all over the globe. Other studies found different kinds of in-group biases of third parties when they decide on punishment both in real groups (Bernhard, Fischbacher & Fehr, 2006; Hoff, Kshetramade & Fehr, 2011) and in experimentally induced minimal groups (Goette, Huffman & Meier, 2012; Butler, Conzo & Leroch, 2015). Finally, Jordan, McAuliffe & Warneken (2014) find that already young children exhibit an in-group bias when inflicting third party punishments.

the notable exception of Bartling et al. (2014), all of these studies did not consider the sanctioned player's intentions as a possible explanation for the decision to punish or reward but have rather focused on the realized outcome.

Erik de Kwaadsteniet et al. (2013) study third party punishments and rewards in a common pool resource setting. Like previous experiments, they examine how third party punishment and reward decisions depend on the realized outcome and find that third parties use the same benchmark for both sanctions: Equality of payoffs. Importantly, they also elicited their participants' experienced emotions in response to the observed behavior and report an interesting finding: Despite relying on the same benchmark, punishments were driven exclusively by negative emotions whereas rewards were driven exclusively by positive emotions. Our results show that decisions to punish and decisions to reward are not just different from this psychological perspective but also differ fundamentally on the level of processing.

In the next section we will describe our experimental design in more detail. Then we derive testable hypotheses from existing theories followed by the presentation of our results in section IV. The paper concludes in a discussion of these results in section V.

II. Experimental Design

Our participants played ten mini dictator games in groups of four. Each group consisted of a dictator, a recipient, and two third parties: One mechanism designer and one mechanism implementer. In each game, the dictator had to decide between two possible allocations of ten points between herself and the recipient. Throughout the paper we will refer to the allocation that yields the higher payoff for the dictator as the selfish choice and the respective other allocation as the nice choice.

Before the dictator chooses an allocation, the mechanism designer decides whether the mechanism implementer will be able to punish (-2 points) or to reward (+2 points) the dictator after she chooses an allocation. The dictator is not informed about the mechanism designer's decision when she chooses between the two allocations however. Finally, the mechanism implementer learns both the decision of the designer and that of the dictator and has to decide whether or not to apply the mechanism provided by the designer. Figure 1 depicts the stage game in extensive form.

The mechanism implementers made their decisions using the strategy method, i.e., they had to decide for each of the four possible outcomes whether they wanted to apply the provided mechanism or not. After the participants had finished making their respective decisions in the mini dictator games, we elicited their beliefs about the behavior of the dictators. More specifically, we showed them each game a second time and asked them to estimate what share of dictators chose which allocation (see procedural details in Appendix A as well as a translation of the instructions in Appendix B).

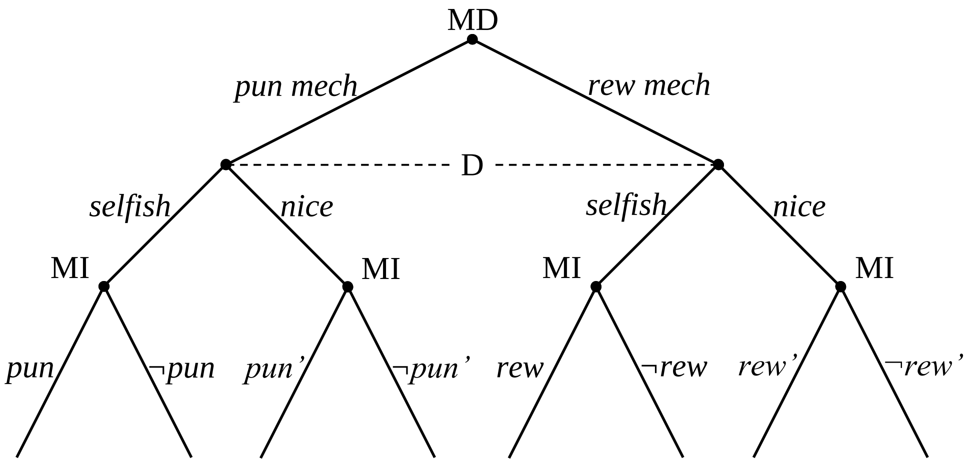


FIGURE 1. THE STAGE GAME IN EXTENSIVE FORM.

Notes: The dashed line indicates the information set around the two decision nodes of the dictator. The decision nodes of the mechanism implementer are marked by "MI" and that of the mechanism designer by "MD". "pun/rew mech" indicates the designer's choice for the punishment/reward mechanism, respectively. "selfish/nice" are the dictator's decisions for the selfish/nice allocation. "pun/¬pun" is the implementer's decision to punish/not punish and "rew/¬rew" accordingly indicate the decisions to reward/not reward.

All games were chosen such that the dictator never had a strategic incentive to be nice: Her payoff from the selfish allocation was never smaller than the total payoff she would have gotten from being nice and getting the reward on top (the payoff structure of all games is depicted in Table 1 below, Table 8 in Appendix C additionally shows descriptive statistics for all choice data). This feature of our design prevents that the true intention behind choosing the nice allocation could become ambiguous such that an implementer should never withhold a reward because she might be in doubt about whether the dictator is in fact nice or merely hoping for the reward.

Further note that we ruled out potential efficiency concerns of the two third parties by introducing a redistribution rule: If a mechanism implementer decided to deduct two points from the dictator, these points were not lost but rather equally redistributed among the dictators and recipients in all other groups. Likewise, if a mechanism implementer rewarded two points to her dictator, these points were deducted from the dictators and recipients from all other groups in equal parts. We also excluded the possibility that mechanism implementers might use the mechanisms in order to discipline the dictators and influence their future decisions by withholding all feedback until the end of the experiment.

TABLE 1—PAYOFF STRUCTURE OF THE TEN GAMES.

Game	1	2	3	4	5	6	7	8	9	10
Nice	7,3	6,4	5,5	6,4	5,5	5,5	4,6	2,8	4,6	3,7
Selfish	9,1	9,1	9,1	8,2	8,2	7,3	8,2	8,2	6,4	5,5

Notes: Each column corresponds to one game. The numbers in row "Nice"/"Selfish" show the payoffs if the nice/selfish allocation is chosen. The first number in each cell corresponds to the dictator's payoff and the second number corresponds to the recipient's payoff.

In order to allow for within subject comparisons, we had two participants in every group decide in both the role of the mechanism designer and the mechanism implementer. At the very end of the experiment it was randomly determined for each group whose decisions in each role were going to be implemented. To control for potential sequence effects, one half of the third parties first decided in the role of the mechanism designer whereas the other half first decided in the role of the mechanism implementer. Also the order of decisions within these two blocks was fully randomized for each participant.

Both the mechanism designer and the implementer received an endowment of 9 points in every game. Whenever an implementer decided to punish or reward the dictator, she had to bear a cost of half a point. There were no costs for the designers.

At the end of the experiment, one of the ten games was randomly selected for payout and the participants earned 1 Euro for each point they had earned in this game. While the experimenter prepared the receipts, the participants filled out a questionnaire for which they received an additional 4 Euros. They were further paid a show-up fee of 2 Euros such that the average earnings amounted to 19.30 Euros for an average duration of 85 minutes.

We conducted four sessions with a total of 104 participants (i.e., 52 third parties). The experiment was programmed and conducted using the software z-Tree (Fischbacher, 2007). Our participants (41% male, average age 20.9 years) were recruited from the student population of the University of Konstanz using ORSEE (Greiner, 2004).

III. Hypotheses

A. *Mechanism Implementers: Behavior*

How should the mechanism implementers come to their punishment and reward decisions? The two main motives for punishments and rewards discussed in the literature are payoff inequalities and the players' intentions behind their actions.

HYPOTHESIS 1. IMPLEMENTER BEHAVIOR

The probability of a punishment decreases and that of a reward increases in

a. *the dictator's transfer to the recipient, and*
b. *the dictator's revealed intention.*

Outcome oriented models of inequity aversion like those of Fehr & Schmidt (1999), Bolton & Ockenfels (2000), or Charness & Rabin (2002) state that people dislike uneven outcomes and predict that the greater the difference between payoffs, the greater the willingness to pay to reduce this difference. Part a. of the above hypothesis directly follows from this logic: Lower transfers are more worthy of punishment and higher transfers more worthy of reward.

The second part of the hypothesis is derived from models that take a player's intention into account like those of Rabin (1993) or Levine (1998). They state that people reciprocate their opponent's kindness, i.e., they are nicer to nice others and meaner to mean others. We will focus on Levine's model here, as the kindness concept of Rabin is redundant for our experimental design.[2]

Levine assumes that people differ in the degree to which they care about others. Technically, they differ in the weight they attribute to the other's payoff in their utility function. This weight can be positive, which means that they are of a kind type that likes to help others, or it can be negative, which implies that they are of an unkind type willing to harm others. With the additional assumption that people also care about the type of their opponents, Levine further includes a motive for reciprocity: People are kinder to others of a kind type and unkinder to others of an unkind type. People's decisions reveal information about their type: For instance, a low offer in the ultimatum game indicates that the proposer has a low or even negative concern for the responder—and therefore deserves rejection. If there is a binary choice between two differently kind options, then the frequency of the unkind choice can be used as a measure of the unkindness. Those who make the unkind choice are the least kind, and the smaller this group, the unkinder they must be on average. Thus, if few people choose the unkind option, one can conclude that those that do are particularly unkind. Similarly, if few people are kind, those who are—are particularly kind.

This means that we can use our participants' beliefs about the dictators' behavior to construct a measure for their expected (un)kindness: Suppose a mechanism implementer believes that 20% of all dictators would choose the selfish option in a given game. Then if she observes a specific dictator to indeed choose the selfish allocation in that game, she can update her belief about this dictator's type to the expected kindness of the population's least kind 20%. Likewise, if

[2] Due to the binary structure of our games, a player's kindness in the sense of Rabin is always equal to -.5 for a dictator that chooses the selfish option and always +.5 for those that choose the nice option.

she observes a dictator to choose the nice allocation, she can update her belief about the dictator's type to the expected kindness of the population's most kind 80%. In our analyses we assume for simplicity that the distribution of types is uniform between -1 and 1. Given the belief of mechanism implementer i about the behavior of all dictators in game k, we can thus easily compute a measure of the expected kindness/revealed intention of dictator d depending on her choice in game k as:

$$Intention_{dki} = \begin{cases} 1 - b_{ik}(nice) & \text{if } d \text{ chose nice} \\ -b_{ik}(nice) & \text{if } d \text{ chose selfish.} \end{cases} \tag{1}$$

Where $b_{ik}(nice)$ is mechanism implementer i's belief about the share of dictators that are nice in game k.

B. Mechanism Designers: Behavior

How should the mechanism designers decide whether punishment or reward is the best mechanism for a given situation? We consider two general motives for the mechanism designer. One efficiency motive (minimize the expected number of implemented rewards and punishments) and one social motive (maximize the fit between the chosen mechanism and the dictator's expected choice).

HYPOTHESIS 2. DESIGNER BEHAVIOR

The probability of a mechanism designer to choose the reward mechanism increases in

a. the expected relative efficiency of the reward mechanism, and in

b. the dictator's expected relative worthiness of reward.

A mechanism designer might want to make sure that the mechanism implementer becomes active only as rarely as possible in order to minimize sanctioning costs to the society. This is equivalent to always choosing the mechanism which will be applied the least. Formally, we measure the expected relative efficiency of the reward mechanism in game k based on designer i's belief about the share of selfish dictators, $b_{ik}(selfi)$, as

$$RelEffRew_{ki} = b_{ik}(selfi)\left(\sigma_{MI,k}(pun\ selfi) - \sigma_{MI,k}(rew\ selfi)\right) + (1 - b_{ik}(selfi))\left(\sigma_{MI,k}(pun\ nice) - \sigma_{MI,k}(rew\ nice)\right), \tag{2}$$

where $\sigma_{MI,k}(m\ a)$ denotes the share of mechanism implementers that apply mechanism m when they face a dictator that chose allocation a in game k. Since applying any mechanism is costly to the implementer, efficiency can be maximized by the designer by choosing the reward mechanism whenever the punishment mechanism will more likely be applied and vice versa. This establishes the first part of hypothesis 2.

The second motive we consider is a social one: Mechanism designers might want to maximize the fit between the chosen mechanism and the dictator's potential worthiness of punishment or reward. For instance, if choosing the selfish allocation is considered as particularly worthy of punishment but choosing the nice allocation is not necessarily worthy of reward, then such a designer should choose the punishment mechanism.

We use three different measures for the dictator's expected relative worthiness of reward. One based on each game's potential outcomes, one based on the dictator's potential intentions, and finally one based on the actual punishing and rewarding decisions of the mechanism *implementers*.

Following the logic of outcome based models of inequity aversion, the mechanism designer should choose the reward mechanism if the dictator's expected transfer is high and the punishment mechanism if it is low. Of course, the expected transfer in game k depends on the individual designer i's belief about the dictators' behavior. Formally, we will compute the expected transfer as

$$E[Transfer_k|b_{ik}] = b_{ik}(selfi)\ \pi_{Rk}(selfi) + \left(1 - b_{ik}(selfi)\right)\pi_{Rk}(nice) \tag{3}$$

Where $\pi_{Rk}(a)$ denotes the recipient's payoff in game k when the dictator chooses allocation a. The expected transfer will be our first measure for the dictator's expected relative worthiness of reward.

Since the mechanism designer decides without knowing the dictator's choice, the expected intention is always equal to zero and thus uninformative. However, a designer could nevertheless take the dictator's potential intention into account by selecting the mechanism that fits the more extreme of the two. If only few dictators are expected to be selfish, then those who are, really deserve a punishment. However, being one of the many other dictators that choose the nice allocation does not make each of them particularly nice such that they do not really deserve a reward. In such a case the designer should choose the punishment mechanism if she wants to make sure that those who really deserve it, can in fact get punished. Put differently, the designer should choose the mechanism which fits those dictators that she expects to be rare. Our second measure for the dictator's expected relative worthiness of reward will thus be the designer's believed share of selfish dictators. The higher it is, the more likely should the designer choose the reward mechanism.

Finally, our third measure for the dictator's expected worthiness of reward is called reward prevalence. We construct it by aggregating the choice data of the mechanism implementers for each of the ten games as shown below. The general idea behind this measure is the following: If one mechanism is implemented more often than the other, then the dictator's potential behavior is more worthy of that mechanism than of the other and then also the designers should supply this mechanism more often. Formally, the reward prevalence in game k is defined as:

$$RP_k = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{ik}(rew\ nice) + \mathbb{1}_{ik}(rew\ selfi) - \mathbb{1}_{ik}(pun\ selfi) - \mathbb{1}_{ik}(pun\ nice). \tag{4}$$

Where $\mathbb{1}_{ik}(rew\ nice)$ is an indicator function that is equal to 1 if mechanism implementer i decides to reward a dictator that chooses the nice allocation in game k and 0 otherwise (analogously for the other three variables). The reward prevalence of a game can thus take on values in the interval from -2 to 2. Positive values indicate that more implementers would reward a dictator in this game than they would punish. We thus expect a positive relationship between the share of designers choosing the reward mechanism for a game and the game's reward prevalence.[3]

C. Decision Times

Psychological evidence-accumulation models as first developed by Roger Ratcliff (1978) make predictions about the decision times in binary choice tasks. Such models state that information is continuously sampled until sufficient evidence in favor of one of the options is accumulated and the option is then chosen. Simply put, they predict that people need more time to decide, the more difficult the decision. In this context, a decision is easy when one alternative spends much more utility than the other. If this is the case, more evidence in favor of the preferred option is accumulated at each point in time such that the respective decision barrier is quickly reached. When the two alternatives become more similar however, coming to a decision becomes more difficult and requires more time. This means that decision times are longer, the closer one gets to indifference. If implementers indeed process outcomes and intentions as suggested by hypothesis 1, then this should also become apparent in their decision times in a specific manner.

HYPOTHESIS 3. IMPLEMENTER DECISION TIMES

Mechanism implementers decide faster for more extreme values of

a. the dictator's revealed intention, and

b. her transfer to the recipient,

while they become slower for more intermediate levels of the two.

[3] Note that, due to our within-subjects design, our measure for reward prevalence can in principle also be constructed for each mechanism designer individually using only his or her own decisions in the role of the mechanism implementer. This individual measure should obviously capture each individual designer's preferences better than the aggregate one. But since it can only take on 5 different values, it can also miss finer differences that could be captured by the aggregate version we introduced above. As it turns out, we obtain virtually identical results with either version of the measure and therefore only report one here.

Consider first the mechanism implementers' decision times depending on the revealed intention of the dictator. According to the model by Levine (1998), every mechanism implementer should have two threshold levels of revealed intention: One punishment threshold and one reward threshold. Only dictators with a revealed intention below the punishment threshold should be punished and only those above the reward threshold should be rewarded. When an implementer faces a dictator who revealed an intention equal to one of these thresholds she will be indifferent between using and not using the respective mechanism. Seen in light of the evidence-accumulation model, this means that decision times should peak at these indifference points and become faster the further away the revealed intentions are from them.

Let us now consider the case of outcomes. Outcome based models of inequity aversion predict that an implementer should punish or reward as soon as the associated costs are smaller than the utility gained from the reduced inequality. This means that, also with respect to the dictator's transfer to the recipient, every mechanism implementer should have two thresholds. If the dictator's payoff is above the punishment threshold she should be punished and if it is below the reward threshold she should be rewarded. As for the case of intentions, the evidence-accumulation model predicts that decision times should peak at the respective thresholds and become faster the further away the realized outcomes are from them.

Analogously for the mechanism designers' decision times, we hypothesize:

HYPOTHESIS 4: DESIGNER DECISION TIMES

Mechanism designers decide faster for more extreme values of

- a. *the expected relative efficiency of the reward mechanism, and*
- b. *the dictator's expected relative worthiness of reward,*

while they become slower for more intermediate levels of the two.

The predictions of evidence accumulation models about the mechanism designers' decision times are straight forward. If the relative efficiency of the two mechanisms is processed as suggested by hypothesis 2, then a decision should become more difficult—and decision times longer—the smaller the differences in expected efficiency between the two.

Similarly, if a designer's decision is based on the dictator's expected relative worthiness of reward, then it should be particularly easy to decide whenever the dictator's expected behavior is particularly worthy of either reward or punishment but not the respective other. To test this part of hypothesis 4, we will use the same three measures of the dictator's expected relative worthiness of reward as for the behavioral hypothesis: The expected transfer to the recipient, the expected share of selfish dictators, and finally the reward prevalence. For more extreme values of these measures we expect faster decision times while they become slower for more intermediate values.

IV. Results

A. Mechanism Implementers: Behavior

Consistent with previous findings, we find that our mechanism implementers are indeed willing to pay in order to punish and/or reward the dictators—only 15% of all implementers followed the standard prediction and never applied any mechanism. Figure 2 shows the share of implementers that rewarded and punished nice and selfish dictators respectively.

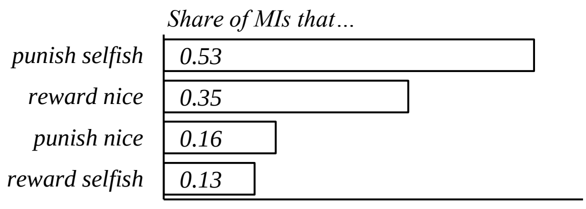


FIGURE 2. SHARE OF IMPLEMENTERS APPLYING THE PROVIDED MECHANISM GIVEN THE DICTATOR'S CHOICE.

Naturally, a decision to reward or punish another's decision depends on the situation at hand. Thus even though we observe significantly more punishments than rewards, these results should not be understood as an indication of a general preference for punishments over rewards. The scatterplots in Figure 3 and 4 nicely illustrate this point. In Figure 3 we see the share of implementers that used the available mechanism in each of the games plotted against the dictator's chosen transfer to the recipient. There is a clear trend: The higher the transfer, the fewer implementers will punish and the more will reward. In Figure 4 the share of implementers that punish and reward are now plotted against the average revealed intention of the dictators.[4] Also here, we can see a trend: The better a dictator's intention, the more implementers reward and the fewer punish.

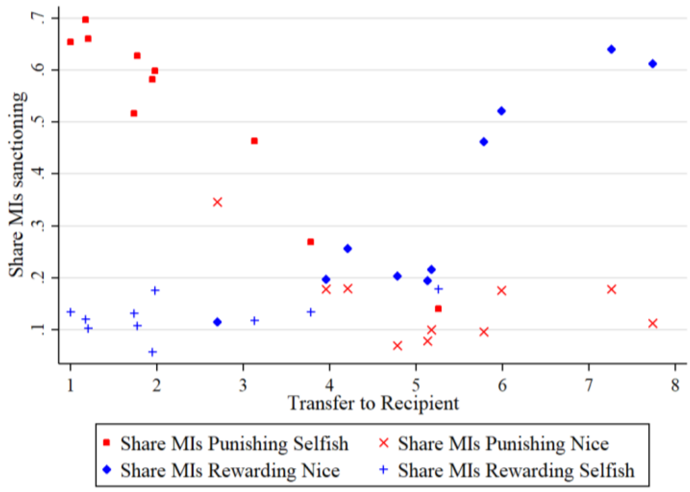


FIGURE 3. SHARE OF IMPLEMENTERS APPLYING THE AVAILABLE MECHANISM IN EACH GAME GIVEN THE DICTATOR'S TRANSFER TO THE RECIPIENT.

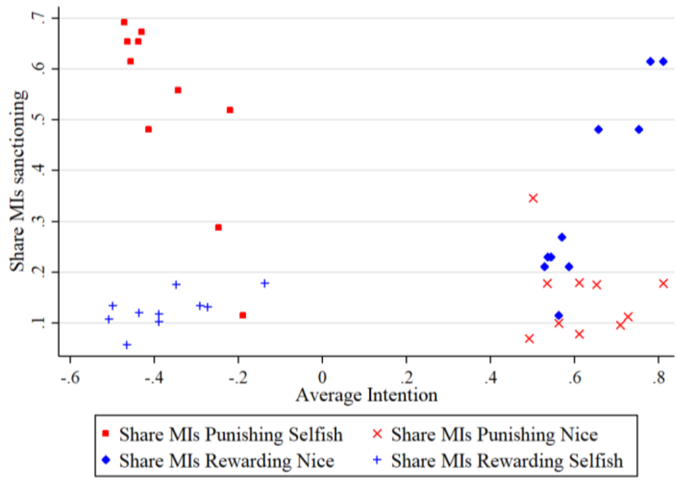


FIGURE 4. SHARE OF IMPLEMENTERS APPLYING THE AVAILABLE MECHANISM IN EACH GAME GIVEN THE DICTATORS' AVERAGE REVEALED INTENTION IN EACH GAME.

[4] The average revealed intention simply corresponds to the intention measure introduced above but using the average belief of all implementers instead of the individual beliefs.

To put these eyeballing results to a statistical test, we ran a series of conditional fixed effects logit regressions with the decisions to punish or to reward as the dependent variables. Note that there is a natural distinction between intuitive and counterintuitive situations for the implementers in our design: Whenever the implementer faces a dictator who chose the selfish (nice) allocation, it is more intuitive to punish (reward) and vice versa. To account for this distinction, we ran the regressions for three subsets of our data: 1. Only intuitive situations, 2. Only counterintuitive situations, and 3. Intuitive and counterintuitive situations pooled together. As will become apparent below, this distinction is particularly relevant on the level of processing. We present the results from the two separate subsets of intuitive and counterintuitive situations in the main part of the paper and those from the pooled dataset in Appendix D.

Table 2 shows the results from the regressions on the intuitive situations. Our behavioral hypothesis about the mechanism implementers is broadly confirmed: The likelihood of being punished as a selfish dictator is significantly decreasing in the dictator's transfer as well as in her revealed intention (models 1 and 2). This also holds when the two explanations are tested jointly (model 3). However, the coefficient on the revealed intention loses its significance in the joint model, indicating that the transfer plays a more important role in determining the punishment decision. This view is corroborated when we compare the Pseudo-R^2 measures of the two single parameter models: The size of the transfer can explain a much larger share of the observed variation than the revealed intention.

Likewise, the likelihood of getting a reward is significantly increasing both in the transfer to the recipient and the dictator's revealed intention (models 4 and 5). Again, both effects also hold qualitatively when tested jointly in model 6 though the coefficient on the dictator's intention again loses its significance. Also here the Pseudo-R^2 measure is much larger in model 4 than in model 5, indicating that also for the reward decisions, the size of the transfer is more important than the revealed intention.[5]

The results for the counterintuitive situations are reported in Table 3. Note first that the values of the Pseudo-R^2 are generally very low in all six models indicating that neither explanation can capture behavior in the counterintuitive situations very well. Model 1 shows that a dictator who chose the nice option is more likely to get punished the smaller the size of the transfer to the recipient. So even when the dictator chose the objectively less selfish option, smaller transfers are still significantly more likely to get punished. In model 2, we find a negative but insignificant effect (p=.36) of the revealed intention on the likelihood of a punishment of a nice choice. When the two explanations are tested jointly in model 3, the effect of the size of the transfer remains negative and highly significant whereas the effect of the intention changes its sign and becomes even less significant (p=.70).

In models 4-6, we find positive but insignificant coefficients on both transfer and intention. Along with the very low Pseudo-R^2's this indicates that neither differences in the size of the transfer nor in the intention can explain well, why some implementers reward dictators that chose the selfish option.

In the models using the pooled data set (see Table 9 in Appendix D) the probability for punishment (reward) is also highly significantly decreasing (increasing) in the size of the transfer. With respect to the revealed intention, we see the same effects in the pooled dataset as in the intuitive situations with the one ex-

[5] In addition to the dictator's transfer to the recipient and her intention as defined above, we also ran regressions in which we included a dummy variable *ChosePunMech*. It is equal to 1 if the respective implementer selected the punishment mechanism for this game when deciding in the role of the designer and 0 otherwise. This allows us to control for the possibility that a subject might prefer a mechanism for a specific game for reasons that are not captured by outcome and intention like a general preference for one of the mechanisms. Even though choosing the punishment mechanism for a game tends to increase (decrease) the likelihood of actually punishing (rewarding) in the same game, we do not report these results here since we obtain virtually identical results for all other parameters.

ception that, in the pooled data, the probability for a punishment is also significantly decreasing in the revealed intention when it is tested jointly with the size of the transfer (model 3).

TABLE 2—CONDITIONAL FIXED EFFECTS LOGITS OF IMPLEMENTER DECISIONS: INTUITIVE SITUATIONS.

Dependent Variable:	Punished Selfish			Rewarded Nice		
	(1)	(2)	(3)	(4)	(5)	(6)
Transfer	-0.994***		-0.941***	0.963***		0.930***
	(0.119)		(0.124)	(0.120)		(0.126)
Intention		-2.436***	-0.839		2.361***	0.476
		(0.509)	(0.591)		(0.520)	(0.593)
Pseudo-R^2	.295	.076	.301	.298	.072	.300
N	400†	400†	400†	370‡	370‡	370‡

Notes: Robust standard errors in parentheses. Fixed effect group variable: Subject. † 12 (‡15) subjects dropped due to all positive or negative outcomes.
***Significant at the 1 percent level.

TABLE 3—CONDITIONAL FIXED EFFECTS LOGITS OF IMPLEMENTER DECISIONS: COUNTERINTUITIVE SITUATIONS.

Dependent Variable:	Punished Nice			Rewarded Selfish		
	(1)	(2)	(3)	(4)	(5)	(6)
Transfer	-0.378***		-0.394***	0.162		0.133
	(0.112)		(0.120)	(0.138)		(0.148)
Intention		-0.548	0.270		0.715	0.440
		(0.598)	(0.688)		(0.793)	(0.845)
Pseudo-R^2	.060	.004	.061	.011	.006	.013
N	280†	280†	280†	190‡	190‡	190‡

Notes: Robust standard errors in parentheses. Fixed effect group variable: Subject. † 24 (‡33) subjects dropped due to all positive or negative outcomes.
***Significant at the 1 percent level.

B. Mechanism Implementers: Decision Times

Table 4 and Table 5 show the results of fixed effect regressions on the mechanism implementers' natural-log decision times in the intuitive and the counterintuitive situations, respectively.[6] In order to test for the hypothesized quadratic effects of the dictator's transfer and revealed intention we also included their respective squared terms. We further included the decision number as an explanatory in all models to control for potential practice effects.

The first three models in Table 4 confirm our hypothesis regarding the decision times for the punishment decisions in intuitive situations: As predicted by the evidence-accumulation model, the first two models show that mechanism

[6] As for the behavioral results, we present the analysis for the pooled data in Appendix D.

implementers are indeed significantly slower to decide on punishment of a selfish choice when (a) the dictator's transfer and (b) the dictator's revealed intention are intermediate. Decision times are estimated to peak around a transfer of 2.7 points and a revealed intention of about -1/2. The further away from these thresholds, the faster the decision times. The R^2 of the first and the second model are almost identical, i.e., the dictator's transfer and intention seem to explain the mechanism implementers' decision times equally well.

Our hypothesis does not find support in our data with respect to the decision times for the reward decisions. Even though we do find highly significant effects of the transfer and its squared term, the effect goes in exactly the opposite direction as expected: Decision times are estimated to reach a *minimum* around a transfer of 5.4 points and become *slower* for higher and lower transfers. This result is puzzling.

The coefficients on the revealed intention and its squared term are insignificant and also their signs are inconclusive: In the joint model they are as hypothesized but turn to the exact opposite when the transfer variables are not included (model 5). It thus seems that, even though both outcomes and intentions are relevant in the intuitive situations both for decisions to reward and for decisions to punish, the motives are processed in a fundamentally different way when deciding about rewards and when deciding about punishments.

The results from the counterintuitive situations shown in Table 5 are similarly inconclusive. The decision times in the counterintuitive situations are not well captured neither by the size of the dictator's transfer to the recipient nor her revealed intention. With just one exception, all coefficients are far from significant. Only when implementers can punish nice dictators (models 1-3), the size of the transfer seems to significantly affect their decision times: We find a negative linear effect and a positive quadratic effect such that decision times are estimated to be *fastest* at a transfer of around 5.9 and become *slower* for more extreme values. We lack a meaningful interpretation of this finding however. A potential explanation for the differences between intuitive and counterintuitive situations would be that participants are confused by the counterintuitive situations. However, this should lead to longer decision times in the counterintuitive situations overall which is not reflected in our data (average decision times in all four situations are between 5.17 and 5.31 seconds with similar standard deviations). Given these mixed results, it is not surprising that the results from the pooled data are not very conclusive either (see Table 10 in Appendix D). In summary, we can state that decisions whether or not to punish or reward seem to be processed differently also when the respective mechanism is the less intuitive one for the given choice of the dictator. Apart from this, our results from the counterintuitive situations are difficult to interpret as they either lack significance or are simply puzzling. We now move to the analysis of the mechanism designers.

TABLE 4—FIXED EFFECTS REGRESSIONS OF IMPLEMENTER DECISION TIMES: INTUITIVE SITUATIONS.

Dependent Variable:	*ln(DT_{MI}(punish selfish))*			*ln(DT_{MI}(reward nice))*		
	(1)	(2)	(3)	(4)	(5)	(6)
Transfer	0.217***		0.187**	-0.322***		-0.324***
	(0.083)		(0.084)	(0.110)		(0.111)
Transfer²	-0.040***		-0.035**	0.030***		0.030***
	(0.014)		(0.014)	(0.010)		(0.010)
Intention		-0.820**	-0.609*		-0.170	0.022
		(0.341)	(0.351)		(0.404)	(0.408)
Intention²		-0.810**	-0.616*		0.202	-0.029
		(0.351)	(0.359)		(0.367)	(0.379)
Decision No.	-0.017***	-0.017***	-0.017***	-0.017***	-0.017***	-0.017***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Constant	1.552***	1.652***	1.494***	2.516***	1.722***	2.517***
	(0.107)	(0.072)	(0.119)	(0.296)	(0.107)	(0.305)
R^2	0.147	0.141	0.152	0.145	0.128	0.145
N	520	520	520	520	520	520

Notes: Robust standard errors in parentheses. Fixed effect group variable: Subject.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

TABLE 5—FIXED EFFECTS REGRESSIONS OF IMPLEMENTER DECISION TIMES: COUNTERINTUITIVE SITUATIONS.

Dependent Variable:	*ln(DT_{MI}(punish nice))*			*ln(DT_{MI}(reward selfish))*		
	(1)	(2)	(3)	(4)	(5)	(6)
Transfer	-0.401***		-0.408***	0.052		0.040
	(0.114)		(0.116)	(0.075)		(0.077)
Transfer²	0.034***		0.035***	-0.014		-0.011
	(0.010)		(0.010)	(0.013)		(0.013)
Intention		0.011	0.137		-0.390	-0.277
		(0.418)	(0.420)		(0.311)	(0.321)
Intention²		-0.055	-0.148		-0.329	-0.260
		(0.379)	(0.390)		(0.319)	(0.327)
Decision No.	-0.019***	-0.020***	-0.019***	-0.017***	-0.017***	-0.017***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Constant	2.853***	1.778***	2.849***	1.734***	1.689***	1.701***
	(0.304)	(0.110)	(0.314)	(0.096)	(0.064)	(0.108)
R^2	0.177	0.154	0.178	0.179	0.176	0.181
N	520	520	520	520	520	520

Notes: Robust standard errors in parentheses. Fixed effect group variable: Subject.
***Significant at the 1 percent level.

C. *Mechanism Designers: Behavior*

Figure 5 shows the distribution of the frequencies with which the designers chose the reward mechanism. On average, they selected the reward mechanism in 40.4% of the cases, but there is large heterogeneity: Not a single subject chose the reward mechanism for all ten games and only one subject never did. All others sometimes chose the punishment mechanism and sometimes the reward mechanism in varying frequencies. We take this as a first indication that they care about the situation at hand instead of having a strong general preference for one or the other mechanism.

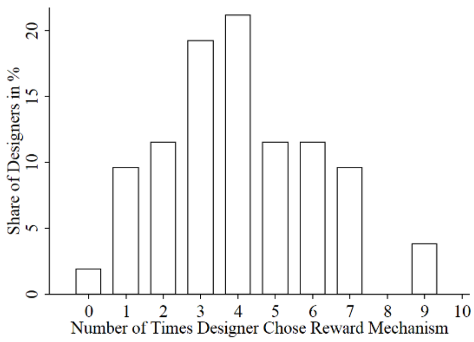


FIGURE 5. DISTRIBUTION OF DESIGNERS CHOOSING THE REWARD MECHANISM.

The scatter plots in Figure 6 show the share of mechanism designers choosing the reward mechanism in each of the ten games. The top left panel clearly shows that mechanism implementers do not take efficiency into account as hypothesized. The share of mechanism implementers that choose the reward mechanism is in fact even *decreasing* in its expected efficiency advantage over the punishment mechanism. In contrast, the effects of the expected transfer (top right), the believed share of selfish dictators (bottom left), as well as the reward prevalence (bottom right) are all positive as hypothesized.

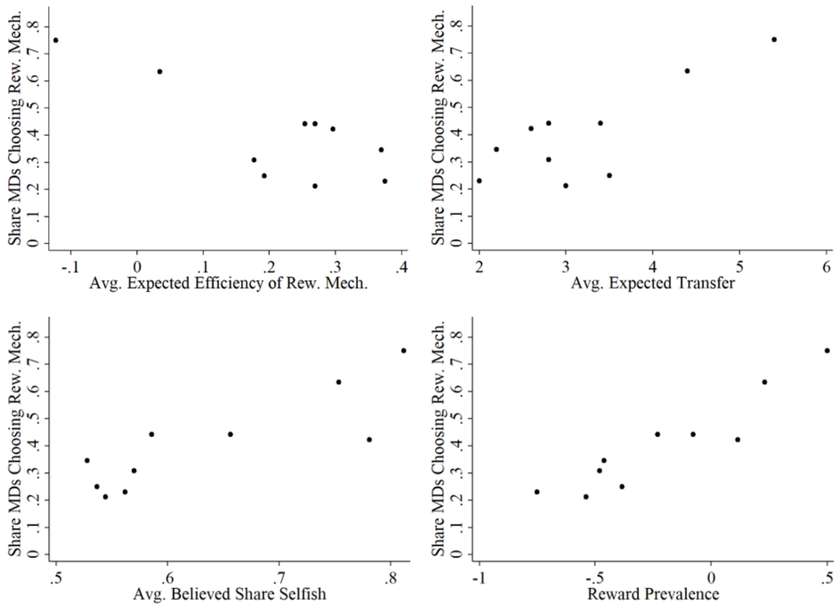


FIGURE 6. SHARE OF MECHANISM DESIGNERS CHOOSING THE REWARD MECHANISM.

Notes: Plotter over: *(top left)* average relative efficiency advantage of rewards over punishments, *(top right)* average expected transfer to the recipient in each game, *(bottom left)* average believed share of dictators choosing the selfish option in each game, *(bottom right)* reward prevalence in each game.

Table 6 reports the results of six fixed effects logit regressions on the mechanism designers' decisions. The first four models show that all four effects visible in figure 4 are highly significant. How do the four explanations compare with respect to their model fit? The reward prevalence yields by far the highest Pseudo-R^2, followed by the expected transfer, the relative efficiency of the reward mechanism, and finally the believed share of selfish dictators.

Model 5 jointly tests the three variables that we use to measure the dictators' potential worthiness of reward or punishment (Transfer, Belief, and Reward Prevalence). We see that all signs remain positive but only the coefficient on the reward prevalence remains significant.[7] Finally, model 6 reveals that again only the coefficient on the reward prevalence remains significant when tested jointly with the relative efficiency of the reward mechanism.

TABLE 6—CONDITIONAL FIXED EFFECTS LOGITS OF MECHANISM DESIGNER DECISIONS.

Dependent Variable:	Chose Reward Mechanism					
	(1)	(2)	(3)	(4)	(5)	(6)
Rel. Eff. Reward	-2.793***					-0.141
	(0.537)					(0.719)
E[Transfer|Belief]		0.451***			0.181	
		(0.0847)			(0.209)	
Belief Selfish			1.603***		0.893	
			(0.447)		(0.895)	
Reward Prevalence				2.056***	1.404*	2.006***
				(0.286)	(0.745)	(0.381)
Pseudo-R²	0.0636	0.0669	0.0293	0.127	0.129	0.127
N	510†	510†	510†	510†	510†	510†

Notes: Robust standard errors in parentheses. Fixed effect group variable: Subject. †1 subject dropped due to all negative outcomes. "Rel. Eff. Reward" denotes the relative efficiency of the reward mechanism as defined above. "E[Transfer|Belief]" denotes the dictator's expected transfer based on the designer's belief about the dictator's behavior. "Belief Selfish" denotes the designer's belief about the share of dictators choosing the selfish option.
***Significant at the 1 percent level.
*Significant at the 10 percent level.

D. Mechanism Designers: Decision Times

Table 7 reports the results of a series of fixed effects regressions on the designers' natural-log decision times. All models include a constant term as well as the decision number to control for potential practice effects. The remaining explanatory variables are the same as for the behavioral analysis above but now also including their respective squared terms. Neither the relative efficiency of the reward mechanism, nor the dictator's expected transfer, nor the expected share of selfish dictators can significantly explain the designers' decision times. The only significant effect is that of the reward prevalence: The coefficient on the squared reward prevalence is significantly negative and the one on the reward prevalence itself is small and insignificant, such that the estimated maxi-

[7]There is an important caveat however: Since all three variables measure the dictator's worthiness of reward/punishment, we are likely to have a problem of collinearity such that it is not obvious how to interpret this result. The substantially larger standard errors in model 5 compared to models 2-4 further strengthen this suspicion.

mum decision time indeed lies at a value close to zero as hypothesized. Comparing the first four models' R^2-values also reveals that the reward prevalence is the best single explanation for the decision times that we have.

The three measures for the dictators' potential worthiness of reward and punishment are jointly tested in model 5. Also here the squared reward prevalence is the only significant explanatory variable. This regression must be interpreted with care however: In addition to the potential problem of collinearity which we already mentioned for the behavioral analysis in footnote 7, the additional explanatory variables also increase the risk of overfitting. The estimated coefficients and standard errors in this model could therefore be biased.

Finally, model 6 jointly tests the effects of the reward prevalence and the relative efficiency of the reward mechanism. Again, decision times are estimated to be longest for values of the reward prevalence close to zero and become faster the more extreme the reward prevalence becomes. The relative efficiency of the reward mechanism still has no significant effect. It thus seems that, rather than contemplating the potential efficiency losses caused by the two mechanisms, the designers base their decisions primarily on the dictators' potential worthiness of punishment/reward.

TABLE 7—FIXED EFFECTS REGRESSIONS OF DESIGNER DECISION TIMES.

Dependent Variable:	$ln(DT_{MD})$					
	(1)	(2)	(3)	(4)	(5)	(6)
Rel. Eff. Reward	-0.134					0.048
	(0.176)					(0.211)
(Rel. Eff. Reward)²	-0.007					0.349
	(0.450)					(0.458)
E[Transfer|Belief]		0.075			-0.081	
		(0.083)			(0.104)	
E[Transfer|Belief]²		-0.007			0.003	
		(0.011)			(0.012)	
Belief Selfish			0.117		-0.058	
			(0.428)		(0.460)	
(Belief Selfish)²			0.014		-0.103	
			(0.388)		(0.420)	
Reward Prevalence				0.0528	0.247	0.101
				(0.078)	(0.204)	(0.102)
(Reward Prevalence)²				-0.453**	-0.424**	-0.474**
				(0.180)	(0.197)	(0.186)
Decision No.	-0.067***	-0.067***	-0.066***	-0.067***	-0.067***	-0.066***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Constant	2.181***	1.999***	2.072***	2.252***	2.611***	2.224***
	(0.060)	(0.149)	(0.113)	(0.059)	(0.359)	(0.064)
R^2	0.121	0.122	0.122	0.143	0.146	0.145
N	520	520	520	520	520	520

Notes: Robust standard errors in parentheses. Fixed effect group variable: Subject.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

V. Conclusion

We use a set of ten mini dictator games to study the processes and preferences behind third party punishments and rewards. In our experiment, a mechanism designer first chooses which kind of sanctioning mechanism (punishment or reward) will be made available to a mechanism implementer who then decides on the implementation of this mechanism after observing the dictator's decision.

We find that mechanism implementers are more likely to punish a selfish choice the lower the transfer and the worse the revealed intention of the dictator. Their decision times indicate that they also process the dictator's transfer and intention as predicted by evidence-accumulation models, i.e., they take the most time for intermediate values where it is more difficult to determine whether the choice criterion is reached or not. Similarly, they are more likely to reward a nice choice, the higher the transfer and the better the revealed intention of the dictator. However, the analysis of the decision times reveals that even though rewards too seem to be based on the dictator's transfer and intention, the underlying processes fundamentally differ from those behind the punishment decisions. Instead of a decision time maximum around intermediate values as predicted by the evidence-accumulation model, we find a significant *minimum* around an intermediate transfer and no effect at all for the revealed intention.

Although we still lack a theory that could explain these results, they are in line with findings reported by de Kwaadsteniet et al. (2013). They investigated the mediating effects of positive and negative emotions on third party punishment and reward decisions in a common pool resource game. They report that, behaviorally, their participants used the same benchmark for both the reward and the punishment decisions—equality of payoffs. However, they also find that punishments were exclusively mediated by negative emotions (but not by positive ones) whereas rewards show exactly the opposite effect, indicating that there is an important difference in the way these decisions are processed. More research is needed to fully understand these differences.[8]

When we consider the less intuitive situations in which the implementer can punish a nice choice or reward a selfish choice, the behavioral results still seem to hold qualitatively. However, except for the negative effect of the transfer size on the probability of a punishment, they all loose significance in these situations. Similarly, we neither find strong results in the implementers' decision times in these counterintuitive situations and the only significant effect is difficult to interpret as it indicates that decisions are *easiest* for intermediate transfer sizes and become more and more difficult toward more extreme transfers.

[8] There are some studies on peer punishments and rewards that report another related finding: Andreoni, Harbaugh & Vesterlund (2003), Sefton, Shupp & Walker (2007), Sutter, Haigner & Kocher (2010),Walker & Halloran (2004) all report that people use punishments in a more intuitive fashion than rewards. For instance, Sutter et al. (2010) found that the likelihood of being punished was increasing in the degree to which one's own contribution to a public good was short of the group's average. However, in the case of rewards they found no such relationship: Even though rewards were more likely when contributions were above the group average, a larger positive deviation did not further increase one's chances of being rewarded.

With respect to the mechanism designers, we find evidence that they only consider the dictators' potential worthiness of punishment or reward and completely disregard the efficiency effects of their decisions. In fact, they even seek the mechanism which will more likely be applied by the implementers such that they practically *minimize* efficiency. Of course, we do not claim that this is what actually motivates their decisions but they do seem to dismiss potential efficiency concerns for the sake of ensuring that those who deserve a specific sanction can actually receive it.

More research is needed to gain a deeper understanding of the decisions of both mechanism designers as well as mechanism implementers. A natural next step would be to investigate whether there is systematic individual heterogeneity in the processes and preferences of third party sanctioning decisions and how these are related to other domains.

Appendix

A. Procedural Details

Before entering the lab, participants were informed that they will be randomly allocated to individual cubicles each equipped with a PC, printed instructions, a set of comprehension questions, and a pencil. Once the participants were at their cubicles they started reading the instructions and answering the comprehension questions. The instructions were written in a neutral language, e.g., labeling the participants as A, B, C, and D instead of mechanism designer, dictator, recipient, and mechanism implementer. Only after each participant had correctly answered all comprehension questions, the experimenter read out a short summary of the instructions to establish common knowledge. After the summary, the experimenter started the experiment and the participants then read a precise description of the exact procedures on screen. In this description they learned how the decision situations will be displayed to them on screen and how they will enter their decisions. We took great care to avoid potential measurement biases by individually randomizing many parts of these procedures.

Participants deciding in the role of the mechanism designer saw the two allocations in the form of two pie charts. In the pie charts, the upper, white segment always represented the share of the dictator and the lower, black segment that of the recipient. Whether the nice allocation was displayed on the left or on the right was determined randomly for each game and participant. Next to the pie charts, the respective payoffs for the dictator and recipient were also shown in point values. Decisions were entered using the keys “F” and “J” on a normal computer keyboard. For each individual participant, the assignment of the keys to the actions (choosing the punishment or the reward mechanism) was fixed throughout the experiment, yet it was randomly counterbalanced between participants.

Before each decision screen, participants first saw an information screen on which they were reminded of the key-action assignment. In order to enter the next decision stage they always had to press both decision keys one after the other. As for the key-action assignment, the order in which they had to press the two keys was fixed for each individual but randomly counterbalanced between participants.

The graphical representation of the two allocations was the same for the mechanism implementers as for the designers, with the same randomization rules. However, they had a slightly different key-action assignment: Whenever they wanted to decide against changing the dictator's income, i.e., neither reward nor punish, they had to press the spacebar. In order to punish, they either had to press "F" or "J" and the corresponding other key in order to reward. The key-action assignments in both roles were matched for each participant. Thus if a participant was randomly selected to punish with "F" and to reward with "J" when in the role of the implementer, then he would also choose the punishment mechanism with "F" and the reward mechanism with "J" when in the role of the designer.

The mechanism implementers additionally saw which of the two allocations the dictator had chosen on their decision screens. This was either always the allocation on the left or the one on the right, which was also determined randomly for each participant. They were further reminded of which mechanism was feasible in the current situation above the two pie charts. This is also the only additional information they were displayed on their information screens before each decision situation. In order to get from the information screen to the decision screen, they too had to press the two decision keys relevant for the next situation. For instance, if a participant was randomly selected to punish with "F" and in the next situation the punishment mechanism was feasible, then she would have to press "F" and the spacebar to proceed. Whether they had to press the spacebar or the other key first was again fixed for each individual participant but randomly counterbalanced across all.

Also the decision screens of the dictators had the same graphical representation of the two allocations as for the two third parties. They chose the left allocation by pressing "F" and the right allocation by pressing "J". The information screen was equivalent to that of the mechanism designers, i.e., they were reminded of the key-action assignment and had to press "F" and then "J" or vice versa to proceed.

All participants were instructed that they had to come to a decision within 90 seconds in each situation or otherwise the computer would make a random decision for them and their income in this situation would be zero. This rule never had to be implemented however.

For the elicitation of the participants' beliefs about the decisions of the dictators, we showed them each of the ten games a second time. One half of the participants then had to enter the percentage of dictators they believed to have chosen the allocation on the left and the other half did the same for the allocation shown on the right. For the third parties, the side for which they had to enter their belief was the same as the one on which they saw the allocation chosen by the dictator when they were in the role of the mechanism implementer. Belief elicitation was incentivized using a quadratic scoring rule which gave them a payoff of $0.5 \times \left(1 - \frac{error_i^2}{100^2}\right)$, where $error_i$ is equal to participant i's stated belief (in percent) minus the true percentage of dictators choosing the respective allocation in that session. All ten beliefs were paid out such that they could earn up to 5 additional Euros in the belief elicitation stage.[9]

B. Instructions, Translated from German Original

Instructions were separated into three distinct parts: 1. a set of printed instructions participants read on their own, 2. a summary read out aloud by the experimenter, and 3. further instructions with procedural details on screen directly

[9] Note that hedging cannot be an issue for the beliefs of the third parties since the decisions of the dictators had no (inescapable) influence on their payoffs.

before the experiment started (e.g., how they make decisions using the keyboard, etc.). For the sake of space, we will only show the written instructions for participants in the role of the mechanism designer/implementer here. Other instructions are available on request. Dashed lines indicate page breaks in the original instructions.

General instructions for the participants

Welcome to this economic experiment. Your decisions and possibly the decisions of other participants in this experiment influence your payoff. It is therefore very important that you read these instructions carefully. ***It is forbidden to communicate with the other participants throughout the entire experiment.*** *We therefore ask you not to speak to each other. If something is not clear to you, please take another look at the instructions. If you should still have questions, please give us a sign. We will come to your cubicle and answer your question in person.*

During the experiment we do not speak of Euros but of points. Your income will thus first be computed in points. Your total score in points from the experiment will be converted into Euros in the end, where

1 Point = 1 Euro

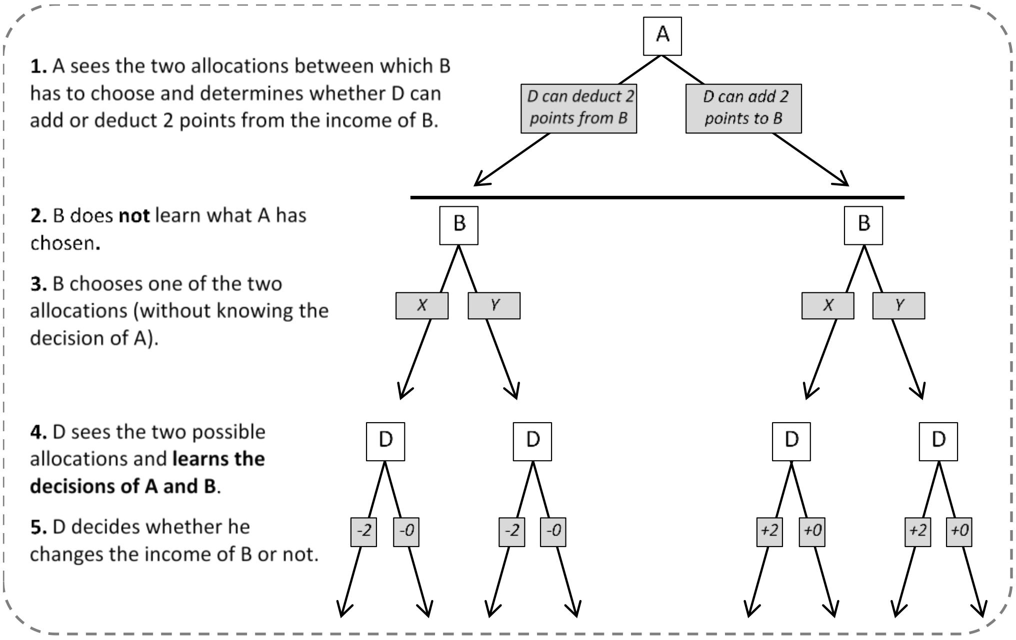
In addition to the points you make during the experiment you earn an additional 4 Euros as a show up fee. On the following pages the experiment will be explained to you.

The experiment at a glance

In this experiment all participants are divided into groups of 4. In each group there is one participant A, one participant B, one participant C, and one participant D. You will learn your own type on the next page.

Participant B chooses one of two possible allocations of 10 points between himself and participant C in 10 different situations. Participant D learns, which of the two allocations participant B has chosen und can then change the income of participant B by 2 points: Whether participant D can deduct 2 points from or add 2 points to the income of participant B is decided by participant A. As opposed to participant D, participant A does not learn which of the two allocations B has chosen but merely what the two possible allocations are. Participant C does not make any decisions.

The sequence for each decision situation is summarized in the following figure:



One of the ten decision situations will be randomly selected for payout at the end of the experiment. You will then receive your income according to your own and the decisions of the other group members in the selected situation. On the following pages we will explain the experiment to you in more detail.

The experiment in detail

Your role

You and one other group member will make decisions both in the role of participant A and in the role of participant D. Only at the end of the experiment chance will determine whether your decisions in the role of participant A or your decisions in the role of participant D will be implemented. Thus, either your decisions as participant A and the decisions of the other group member as participant D or your decisions as participant D and the decisions of the other group member as participant A will be implemented. In contrast, the roles for the other two group members are already fixed. So only one group member will decide in the role of participant B.

The decisions of participant A

For the case that at the end chance decides that you are participant A, you decide for each of the ten decision situations of participant B how participant D can change the income of participant B. You will see one of the ten decision situations of B on your screen at a time and you decide whether D shall be able to deduct or add 2 points to the income of B in this situation. However, you do not know how participant B decides but merely see the two allocation possibilities between which B must decide. You have 90 seconds for each decision. If you do not decide within the 90 seconds, you will receive no income in this situation and chance will decide how participant D can change the income of B.

The income of participant A

If you decide within the 90 seconds, your income as participant A in each decision situation amounts to 9 points. If you do not decide, your income amounts to 0 points.

The decisions of participant B

In each of the ten situations participant B chooses one of the two possible allocations of the 10 points between himself and participant C. In each decision situation the preliminary incomes of participant B and participant C are then equal to the allocation chosen by B. When participant B makes his decisions he does not know whether participant D will be able to deduct to points from or add 2 points to his income. This means that participant B does not know

---

how participant A decided in this situation. Participant B will neither be informed if or how participant D has changed his income. Only after he decided for each of the ten situations he will learn how participants A and D have decided in the situation which was randomly selected for payout. There is no time limit for the decisions of participant B.

The income of participants B and C

In each decision situation the preliminary income of participant B can be changed by 2 points according to the decisions of participants A and D: If participant A decided that participant D can deduct 2 points from the income of B, participant D could accordingly lower the income of B in this situation by 2 points. If participant A decided that participant D can add 2 points to the income of B, participant D could accordingly increase the income of B in this situation by 2 points.

Redistribution of the income changes

If participant D should deduct 2 points from the income of participant B, these points are not lost but are evenly distributed to all participants B and C in all other groups. The sum of incomes of all participants B and C therefore does not change. Also if participant D should add 2 points to the income of participant B, the sum of incomes of all participants B and C cannot change: The added 2 points are evenly deducted from the incomes of participants B and C in all other groups.

This redistribution rule holds for all groups. This means that also the income of participants B and C in your own group depends on the decisions of participants D in all other groups: Every 2 points that are deducted from B's income in some group are evenly redistributed to participants B and C in all other groups and every 2 points that are added to B's income in some group are evenly deducted from the incomes of participants' B and C in all other groups.

The decisions of participant D

For the case that at the end chance decides that you are participant D, you decide for each situation whether you change the income of participant B by 2 points or not. For this, you will see on the one hand the two possible allocations and

---

which of them participant B had chosen and on the other hand whether you can deduct 2 points from or add 2 points to the income of participant B according to the decision of participant A. From you point of view as participant D, there are four possible cases for each of the ten decision situations:

1. *You can* deduct *2 points from B and B chooses the allocation on the* left
2. *You can* deduct *2 points from B and B chooses the allocation on the* right
3. *You can* add *2 points to B and B chooses the allocation on the* left
4. *You can* add *2 points to B and B chooses the allocation on the* right

You decide for each case separately and thus make a total of 40 decisions as participant D. For each decision you have 90 seconds of time. If you do not decide within the 90 seconds, you receive no income in this situation and chance decides if the income of participant B will be changed.

The income of participant D

If you decide within the 90 seconds your income as participant D amounts to either 9 or 8.5 points. If you do not change the income of B, you receive

9 points and if you add 2 points to or deduct 2 points from B's income, you receive 8.5 points. If you do not decide, your income amounts to 0 points.

<u>Summary</u>

Once chance has determined at the end of the experiment who is participant A and who is participant D, the decisions of the randomly determined participant A are implemented. Next the decisions of B are implemented such that for each situation it is clear which of the four cases of participant D has occurred. Finally, the respective decisions of the randomly determined participant D are implemented such that for each situation the income of all participants is fixed. One of the ten situations is then randomly selected and the respective incomes will be paid out to the participants at the end of the experiment. Before the payment there will be two short questionnaires for which you will be paid separately.

You can now start answering the attached comprehension questions. Your answers to these questions have no impact on your payment and exclusively serve your comprehension.

C. Descriptive Statistics

TABLE 8—PAYOFF STRUCTURE AND DESCRIPTIVE STATISTICS OF THE TEN GAMES.

Game	Nice π_D, π_R	Selfish π_D, π_R	% Nice Dictators	Believed % Nice	% MIs Rew. Nice	% MIs Pun. Sel.	% MIs Rew. Sel.	% MIs Pun. Nice	% MDs Chose Pun.
1	7,3	9,1	38.5	43.8	11.5	65.4	13.5	34.6	76.9
2	6,4	9,1	34.6	43.0	26.9	67.3	11.5	19.2	69.2
3	5,5	9,1	42.3	47.2	21.2	69.2	11.5	9.6	65.4
4	6,4	8,2	34.6	45.6	23.1	61.5	5.8	21.2	78.8
5	5,5	8,2	42.3	46.3	23.1	65.4	13.5	9.6	75.0
6	5,5	7,3	46.2	41.4	21.2	48.1	13.5	9.6	55.8
7	4,6	8,2	23.1	34.3	48.1	55.8	13.5	13.5	55.8
8	2,8	8,2	7.7	21.9	61.5	51.9	13.5	11.5	57.7
9	4,6	6,4	19.2	24.7	48.1	28.8	15.4	11.5	36.5
10	3,7	5,5	0	18.8	61.5	11.5	15.4	15.4	25.0

Notes: Columns 2 and 3 show the payoff structure of each game: Column 2 shows the payoffs if the dictator chooses the nice option and column 3 those if she chooses the selfish option. In both columns the first number always represents the payoff of the dictator and the second number that of the recipient. The column "% Nice Dictators" shows the share of dictators that chose the nice allocation in the respective game in percent. "Believed % Nice" is the third parties' average belief about the percentage of nice dictators. "% MIs Rew. Nice" ("% MIs Rew. Sel.") is the percentage of implementers that rewarded a nice (selfish) choice. Analogously, "% MIs Pun. Sel." ("% MIs Pun. Nice") is the percentage of implementers that punished a selfish (nice) choice. Finally, "% MDs Chose Pun." is the percentage of mechanism designers that chose the punishment mechanism for the respective game.

D. Mechanism Implementers: Pooled Data

Table 9 shows the results of the conditional fixed effect logit models for the pooled data. They are basically a mirror image of the behavioral results from the intuitive situations: Higher transfers and better intentions significantly increase the likelihood of a reward and significantly decrease the likelihood of a punishment. Also here a comparison of the single parameter models' Pseudo-R^2's shows that transfers play a bigger role for these decisions than intentions.

TABLE 9— CONDITIONAL FIXED EFFECTS LOGITS OF IMPLEMENTER DECISIONS: POOLED DATA.

Dependent Variable:	Punished (Nice or Selfish)			Rewarded (Nice or Selfish)		
	(1)	(2)	(3)	(4)	(5)	(6)
Transfer	-0.820***		-0.665***	0.695***		0.604***
	(0.061)		(0.081)	(0.0629)		(0.0845)
Intention		-2.308***	-0.694***		2.031***	0.455
		(0.181)	(0.255)		(0.208)	(0.290)
Pseudo-R^2	.334	.253	.343	.280	.195	.283
N	820†	820†	820†	780‡	780‡	780‡

Notes: Robust standard errors in parentheses. Fixed effect group variable: Subject. † 11 (‡13) subjects dropped due to all positive or negative outcomes.
***Significant at the 1 percent level.

Table 10 shows the results of the fixed effects regressions on the mechanism implementers' decision times for the pooled sample. In light of the mixed results from the intuitive and counterintuitive situations, it is not surprising that the regressions with the pooled data do not lend much insight. In punishment situations, the coefficients on the size of the transfer are always far from significant and only the linear coefficient on the dictator's revealed intention is significantly negative in model 2 (but not in the joint model 3). With respect to decision times in all reward situations the puzzling effect we have seen from the intuitive situations prevails: Subjects reward fastest at an intermediate size of the transfer and become slower for more extreme ones. Given that the coefficients from the counterintuitive situations were all insignificant and even had exactly opposed signs we can conclude that this is driven by just the intuitive situations however.

TABLE 10—FIXED EFFECTS REGRESSIONS OF IMPLEMENTER DECISION TIMES: POOLED DATA.

Dependent Variable:	*ln(DT_{MI}(punish))*			*ln(DT_{MI}(reward))*		
	(1)	(2)	(3)	(4)	(5)	(6)
Transfer	-0.029		-0.030	-0.084***		-0.088**
	(0.032)		(0.037)	(0.030)		(0.034)
Transfer²	0.001		0.001	0.009**		0.010**
	(0.004)		(0.004)	(0.004)		(0.004)
Intention		-0.063**	-0.005		-0.040	0.001
		(0.032)	(0.049)		(0.030)	(0.046)
Intention²		-0.035	-0.033		-0.012	-0.043
		(0.052)	(0.054)		(0.049)	(0.050)
Decision No.	-0.018***	-0.018***	-0.018***	-0.017***	-0.017***	-0.017***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	1.865***	1.790***	1.872***	1.887***	1.741***	1.902***
	(0.065)	(0.037)	(0.081)	(0.062)	(0.034)	(0.076)
R^2	0.150	0.148	0.150	0.152	0.147	0.153
N	1’040	1’040	1’040	1’040	1’040	1’040

Notes: Robust standard errors in parentheses. Fixed effect group variable: Subject.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

References

Almenberg, J., Dreber, A., Apicella, C., & Rand, D. G. (2011). Third party reward and punishment: group size, efficiency and public goods. In N. M. Palmetti & J. P. Russo (Eds.), *Psychology of Punishment* (pp. 73-92): Nova Science Publishers.

Andreoni, J., Harbaugh, W. T., & Vesterlund, L. (2003). The carrot or the stick: Rewards, punishments and cooperation. *American Economic Review, 93*(3), 893-902.

Bartling, B., Engl, F., & Weber, R. A. (2014). Does willful ignorance deflect punishment?–An experimental study. *European Economic Review, 70*, 512-524.

Bendor, J., & Swistak, P. (2001). The evolution of norms. *American Journal of Sociology, 106*(6), 1493-1545.

Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. *Nature, 442*, 912-915.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review, 90*(1), 166-193.

Buckholtz, J. W., & Marois, R. (2012). The roots of modern justice: Cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience, 15*(5), 655-661.

Butler, J. V., Conzo, P., & Leroch, M. (2015). Social Identity and Punishment. *GSME Discussion Paper Series, No. 1512*.

Charness, G., Cobo-Reyes, R., & Jiménez, N. (2008). An investment game with third-party intervention. *Journal of Economic Behavior & Organization, 68*(1), 18-28.

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

de Kwaadsteniet, E. W., Rijkhoff, S. A., & van Dijk, E. (2013). Equality as a benchmark for third-party punishment and reward: The moderating role of uncertainty in social dilemmas. *Organizational Behavior and Human Decision Processes, 120*(2), 251-259.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*(2), 63-87.

Fehr, E., & Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives, 14*(3), 159-181.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*, 137-140.

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

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

Gächter, S., Renner, E., & Sefton, M. (2008). The long-run benefits of punishment. *Science, 322*, 1510.

Goette, L., Huffman, D., & Meier, S. (2012). The impact of social ties on group interactions: Evidence from minimal groups and randomly assigned real groups. *American Economic Journal: Microeconomics, 4*(1), 101-115.

Greiner, B. (2004). The online recruitment system orsee 2.0–A guide for the organization of experiments in economics. *University of Cologne, Working Paper Series in Economics, 10*(23), 63-104.

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., & Henrich, N. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science, 327*, 1480-1484.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., & Henrich, N. (2006). Costly punishment across human societies. *Science, 312*(5781), 1767-1770.

Hoff, K., Kshetramade, M., & Fehr, E. (2011). Caste and punishment: The legacy of caste culture in norm enforcement. *The Economic Journal, 121*(556), F449-F475.

Jordan, J. J., McAuliffe, K., & Warneken, F. (2014). Development of in-group favoritism in children's third-party punishment of selfishness. *Proceedings of the National Academy of Sciences, 111*(35), 12710-12715.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review, 76*(4), 728-741.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics, 1*(3), 593-622.

Marlowe, F. W., Berbesque, J. C., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Ensminger, J., Gurven, M., Gwako, E., & Henrich, J. (2008). More 'altruistic' punishment in larger societies. *Proceedings of the Royal Society of London B: Biological Sciences, 275*(1634), 587-592.

Milinski, M., & Rockenbach, B. (2012). On the interaction of the stick and the carrot in social dilemmas. *Journal of Theoretical Biology, 299*, 139-143.

Nikiforakis, N., & Mitchell, H. (2014). Mixing the carrots with the sticks: Third party punishment and reward. *Experimental Economics, 17*(1), 1-23.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review, 83*(5), 1281-1302.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59-108.

Sefton, M., Shupp, R., & Walker, J. M. (2007). The effect of rewards and sanctions in provision of public goods. *Economic Inquiry, 45*(4), 671-690.

Sutter, M., Haigner, S., & Kocher, M. G. (2010). Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *Review of Economic Studies, 77*(4), 1540-1566.

Turillo, C. J., Folger, R., Lavelle, J. J., Umphress, E. E., & Gee, J. O. (2002). Is virtue its own reward? Self-sacrificial decisions for the sake of fairness.

Organizational Behavior and Human Decision Processes, 89(1), 839-865.

Walker, J. M., & Halloran, M. A. (2004). Rewards and sanctions and the provision of public goods in one-shot settings. *Experimental Economics, 7*(3), 235-247.

In Vino Veritas

By Jan Hausfeld & Konstantin Hesler

We study the effect of social drinking on honesty in a lab-in-the-field experiment. Bar-goers are recruited before they enter a bar and—depending on treatment—either participate in the experiment right away (sober condition) or when they come back out from the bar (intoxicated condition). In both conditions, participants' degree of intoxication is measured with a breathalyzer before and after the bar visit. This design allows us to disentangle the effect of acute alcohol intoxication from that of general drinking habits by using the randomly imposed treatment condition as an instrumental variable for the degree of intoxication during the experiment. At the same time, the two breathalyzer measurements are used as a proxy for general drinking habits. We find that people become more honest, the more they had to drink, while at the same time people who drink more in general, are also less honest in general.

JEL Codes: *C93, D03, D63, I10*

Keywords: *Field experiment, honesty, alcohol*

I. Introduction

Proverbs from many different cultures suggest that people become more honest under the influence of alcohol than when sober. Also mere introspection and anecdotal evidence seem to lend support to this conjecture. However, to the best of our knowledge, there is no scientific evidence that can back this claim. We fill this gap by investigating the relationship between social drinking and dishonesty with a lab-in-the-field experiment: We recruited bar-goers to participate in an experiment and measured their level of alcohol intoxication as well as their degree of dishonesty in a simple task. We have two main findings: On the one hand, people indeed become *more* honest, the more they had to drink, while on the other hand, people who drink more in general, are *less* honest in general.

We use a new honesty game in which players first choose between two urns and then privately draw from the chosen urn. To determine their earnings, players then have to report what they have drawn but can be dishonest about it. The game shares the crucial feature of the paradigm introduced by Fischbacher & Föllmi-Heusi (2013): The experimenters cannot verify any individual report, yet they know the theoretical distribution of reports which would be expected under complete honesty. Participants' truthfulness can thus still be assessed by comparing the observed distribution of reports with the one that would be expected if everyone was honest. People that aim to maximize their monetary payoff should always lie to obtain the highest possible outcome. Although many people lied, they often did not take full advantage of the opportunity to lie and a substantial share of people did not lie at all.

Methodologically, our study is most closely related to Morewedge, Krishnamurti & Ariely (2014, study 1) who also recruited bar-goers to participate in an experiment in a mobile lab.[1] They studied the effect of alcohol on behavior in ultimatum games and find that rejection rates increased with stronger intoxication. Interestingly however, offers made by sober and intoxicated participants did not differ in their generosity. The main advantage of such lab-in-the-field experiments is that the real conditions under which alcohol might affect behavior are approximated very closely. This is not the case with lab experiments because participants are not in a natural drinking environment and do not self-regulate their intake. This is also reflected in the higher levels of intoxication reached in the field compared to those induced in typical lab studies.

[1] Bègue et al. (2013) and Duke & Bègue (2015) also ran experiments with bar-goers but conducted the experiments inside the barroom. Bègue et al. (2013) find that subjects' self-perceived attractiveness is increasing in the amount of alcohol they consumed and Duke & Bègue (2015) find that intoxicated participants are more likely to state they would sacrifice the life of one person in order to save five others in a standard trolley dilemma.

There are two straight forward ways in which an effect of alcohol intoxication on honesty in a field experiment could be explained: Either, people behave differently because of their acute degree of intoxication, or it is because people who differ in their drinking habits also differ in their honesty preferences (regardless of their current level of intoxication). To be able to disentangle the two explanations, we introduce treatments to our design. The crucial difference lies in the timing of the experiment. All subjects were recruited before they entered the bar but participated in the main part of the experiment at different points in time. In the *sober* condition, the experiment took place right after recruitment and before subjects went inside. In the *intoxicated* condition however, subjects first entered the bar and only participated after they came back outside to leave. The conditions were randomly imposed and we thus get a clean instrumental variable for the degree of intoxication which cannot be correlated with our participants' general drinking habits.[2] With this approach, we can show that honesty increases in acute alcohol intoxication.

To estimate the effect of general drinking habits we can use the breathalyzer measurements directly: A participant's breath alcohol content measured before and after the bar visit is of course highly correlated with her general drinking habits. Someone who drinks more in general, is also more likely to drink more on the night of the experiment. We find that heavier drinkers are less honest than others who drink less.

The remainder of this paper is structured as follows: We review the related literature in the next section and give a detailed explanation of our experimental design in section III. Our analyses and results are presented in section IV and we conclude in section V.

[2] We rule out the remaining confound that people in the *intoxicated* condition are also more tired than in the *sober* condition by controlling for the time of participation in our regressions.

II. Related Literature

The findings on honesty by Fischbacher & Föllmi-Heusi (2013) have been replicated in numerous other studies (see, e.g., Kajackaite & Gneezy, 2015, Weisel & Shalvi, 2015, or Abeler, Nosenzo & Raymond, 2016, and the respective citations therein) and have also been found to correlate well with dishonesty outside the lab (Hanna & Wang, 2013; Cohn, Maréchal & Noll, 2015; Gächter & Schulz, 2016; Potters & Stoop, 2016; Cohn & Maréchal, 2017; Dai, Galeotti & Villeval, 2017).[3]

Several studies have investigated different determinants of dishonesty. Among the first manipulations were those varying the strength of the incentive to lie (Gneezy, 2005; Mazar et al., 2008; Sutter, 2009; Fischbacher & Föllmi-Heusi, 2013; Kajackaite & Gneezy, 2015) which typically find no or only weak reactions to increased stake size. Some studies explored the relationship between dishonesty and altruism: Cappelen, Sørensen & Tungodden (2013) find that altruism correlates positively with honesty. Biziou-van-Pol et al. (2015) find people who are highly motivated by altruistic preferences to lie less when it is beneficial to themselves and others (Pareto white lies) but to lie more if it is beneficial only to others and costly to themselves (altruistic white lies). Levine & Schweitzer (2015) find that people who were observed to lie prosocially, received higher transfers in a subsequent trust game. Other studies focused on the role of the consequences of a lie, e.g., whether a lie is beneficial or

---

[3] Note that, even though the paradigm by Fischbacher & Föllmi-Heusi (2013) is the most common one, there also are other widely used paradigms in the literature: One is the sender-receiver game introduced by Gneezy (2005) in which a sender learns the true state and sends a (possibly dishonest) message about the state to the receiver. The receiver then chooses an action and payoffs are determined by this action and the true state. Since the experimenter also knows the true state, truthfulness can be evaluated on an individual basis. In the paradigm introduced by Charness & Dufwenberg (2006) one player can send a promise about a future action to a second player. Different from Gneezy's design, the sender's honesty is then also revealed to the second player (not just to the experimenter). Finally, Mazar, Amir & Ariely (2008) introduced a design in which participants can lie about an outcome that only they observe. In contrast to Fischbacher & Föllmi-Heusi (2013), however, this outcome is the performance in a real effort task. Arguably, this will seem more natural to participants. However, testing becomes more difficult with this design as the true distribution under complete honesty is unknown.

harmful for oneself and/or others (Gneezy, 2005; Erat & Gneezy, 2012; Levine & Schweitzer, 2014; Shalvi & De Dreu, 2014). They find that people care about the externalities of their lies and tend to lie less, the more harm lying does to others. Another strand of literature addresses the question whether people lie more or less when they decide as part of a group rather than individually (Cohen et al., 2009; Wiltermuth, 2011; Conrads et al., 2013; Gino, Ayal & Ariely, 2013; Kocher, Schudy & Spantig, 2017). Here the typical finding is that groups lie more than individuals. Finally, there are several studies exploring the role of gender for honesty preferences, some of which report more lying among males (Dreber & Johannesson, 2008; Friesen & Gangadharan, 2012; Muehlheusser, Roider & Wallmeier, 2015), whereas others find no clear gender effect (Childs, 2012; Gylfason, Arnardottir & Kristinsson, 2013; Arbel et al., 2014; Lohse & Qari, 2014).

Even though honesty and alcohol intoxication have not yet been linked directly, there are two common aspects in both of these areas of research: Cognitive ability and morality. Either one has been studied in both the context of honesty and in that of alcohol intoxication. Cognitive ability is widely known to suffer from alcohol intoxication (see, for instance, Moskowitz, Burns & Williams, 1985; Peterson et al., 1990; Fillmore, Carscadden & Vogel-Sprott, 1998; Curtin & Fairchild, 2003; Schreiber Compo et al., 2011; Morewedge et al., 2014). Findings on the relationship between general cognitive ability and honesty are mixed: Ruffle & Tobol (2017) report a significantly positive effect, Gino & Ariely (2012) find no significant effect and Fosgaard, Hansen & Piovesan (2013) find a significantly negative effect.[4] Further, studies investigating the effect of cognitive load find that subjects under higher load become more

[4] Although these three studies also used different measures for cognitive ability (e.g., Ruffle & Tobol, 2017, used scores on military entrance tests and the short Raven test (Arthur & Day, 1994), whereas Gino & Ariely, 2012, used a test to measure verbal intelligence), all three of them included the cognitive reflection test by Frederick (2005) as one measure for cognitive ability. Another related finding is reported by Capraro (2017b) who conducted a meta-analysis with data from 50 treatments of deception games as introduced by Gneezy (2005). He found less educated people to lie more than others with higher education.

honest than under low load (van 't Veer, Stel & van Beest, 2014; Mann et al., 2015). Similarly, Capraro (2017a) also found time pressure to increase honesty.

These findings indicate that lying is an effortful and more controlled process than being honest. This interpretation is also in line with the results of a brain study on dishonesty by Greene & Paxton (2009) who found increased activity in control related brain regions when participants lied. Two other studies investigating the effect of time pressure on honesty (Gunia et al., 2012; Shalvi, Eldar & Bereby-Meyer, 2012) actually found exactly the opposite effect, i.e., more dishonesty under time pressure. However, these studies share the problem that participants actually knew their payoff maximizing strategies *before* the time manipulation. For instance, in Shalvi et al. (2012), participants had to roll a die as in Fischbacher & Föllmi-Heusi (2013) and only then had to report on the outcome. Both the rolling and the subsequent reporting had to be done either with or without a time limit of 20 seconds. The time limit was started by the participants themselves after having read all instructions. As discussed by Foerster et al. (2013), participants thus could have easily decided on their report before the time limit was even started, shedding doubt on the validity of the manipulation.

Mazar et al. (2008) put attention to moral standards into context with cheating. They find a decrease in dishonesty when participants are reminded of general moral standards at the beginning of the experiment. A similar result is reported by Pruckner & Sausgruber (2013): Moral reminders at unmonitored newspaper sales booths increased payment honesty. In a different strand of literature, alcohol was found to lower one's moral standards: For instance, Denton & Krebs (1990) find that intoxicated participants score lower in a moral maturity test and rate driving under the influence of alcohol as less wrongful (also see MacDonald, Zanna & Fong, 1995). Similarly, Felson et al. (2008) report evidence that people are more likely to commit certain delinquencies like violence, vandalism, car theft, or graffiti-writing under the influence of alcohol.

There is a growing literature exploring the effects of alcohol consumption on different aspects of economic decision making. Most of it has focused on alcohol's effects on risk taking.[5] Burghart et al. (2013) and Proestakis et al. (2013) ran field experiments and both groups found significant effects of breath alcohol content (BAC) on women's (but not men's) risk aversion—however going in opposite directions: Women in the sample of Burghart et al. (2013) become less risk averse with higher BAC, whereas those in Proestakis et al. (2013) become more risk averse. Also the results from different lab studies are inconclusive: Lane et al. (2004) find intoxicated subjects to be less risk averse, Corazzini et al. (2014) find that intoxicated women become more risk averse (no effect for men), Breslin et al. (1999) find no effects at all, and Sjöberg (1969) finds mixed effects depending on the administered dosage. Finally, there are some studies that studied the relationship between general drinking habits and risk aversion. Barsky et al. (1997) and Anderson & Mellor (2008) find people who generally drink more, are also less risk averse (Galizzi & Miraldo, 2017, find the same relationship for women but no effect for men).

III. Experimental Design

The main part of our experiment consists of five rounds of a new honesty game. In each round, subjects first select one of two binary lotteries. The chosen

[5] There also are some studies investigating the effects of alcohol on other domains of economic decision making. Schilbach (2015) found that reducing chronic alcohol consumption increased participants' self-control in a field experiment in India. Corazzini, Filippin & Vanin (2014) and Bregu et al. (2017) studied the effects of alcohol intoxication in the lab on behavior in several standard economic experiments. Bregu et al. (2017) found alcohol to increase altruism whereas Corazzini et al. (2014) find the opposite. Both measured altruism in terms of transfer size in dictator games, but in Corazzini et al. (2014) the recipients were two NGOs instead of other participants as in Bregu et al. (2017). Cooperation rates in strategic games did not significantly differ between the treated and the control group in Bregu et al. (2017). Similarly, Hopthrow et al. (2007) found no effect of alcohol on cooperation rates in prisoners' dilemmas when they were played by individuals but significantly less cooperative choices when they were played in groups. With respect to time preferences, Corazzini et al. (2014) report increased impatience for intoxicated participants. Finally, Burghart, Glimcher & Lazzaro (2013) investigated whether alcohol intoxication changes the degree of violations of the generalized axiom of revealed preferences (GARP) and the independence axiom. While they found slightly more violations of the independence axiom, the GARP turned out to be robust against intoxication.

lottery is then played out, but the true outcome is only observed by the subject. The subject has to report the outcome to determine their income, but can choose to be dishonest about it. The experimenter only observes the outcome reported by the subject.

We ran the experiment in a mobile lab, set up in a van parked in front of bars. People who came to enter the bar were invited to participate in a brief experiment on risk (see Appendix A for more procedural details). Each participant's level of alcohol intoxication was measured twice with a breathalyzer—once before they entered the bar and once after they left.[6]

There were two treatments which differed only with respect to when the actual experiment was conducted. In the *sober* condition, subjects participated in the experiment before they entered the bar, while in the *intoxicated* condition, they participated after they came back out to leave. The treatments were run on different days such that either all or none of the participants inside the bar had already experienced the honesty game. Note that, for the participants in the *sober* condition the first breath alcohol content (BAC) measurement is the one relevant for the experiment whereas the second BAC measurement is the relevant one for those in the *intoxicated* condition. We will use this terminology throughout the paper, i.e., when we refer to the "relevant BAC", this corresponds to the first measurement for subjects in the *sober* condition but to the second measurement for subjects in the *intoxicated* condition.

In addition to the honesty game, we also elicited our participants' level of risk-aversion using a single item coin-flip task as in Gneezy & Potters (1997): Participants received 10 points to bet on a coin-flip. The points they did not bet they kept for sure. For their bet, they received 2.5 times the points they bet if their side came up but got nothing back if the other side came up. We further included a brief test for cognitive ability and a short questionnaire on morality

[6] We used the breathalyzer *AlcoQuant 6020 plus* produced by EnviteC. It has a maximum measurement error of 5 per cent for levels of intoxication below 2‰ (our full range of measurements lies below this threshold) and fulfills the European standard EN 15964 for breath alcohol test devices designed for law enforcement.

and socio-economic background. The test for cognitive ability consisted of ten mathematical equations for which the participants had to indicate whether they were true or false. They had 5 seconds for each equation. The task was not incentivized but participants received a lump sum payment of 1 Euro for working on the task. Our measure for cognitive ability simply corresponds to the number of correctly solved equations. Similarly, our measure for morality is based on the answers to five 11-point Likert-scaled questions like "I would stop at nothing to reach my goals".[7] In order to avoid potential income effects, also participants in the *sober* condition only received feedback on their final payment after they left the bar.

The experiment was programmed using the software z-Tree (Fischbacher, 2007) and conducted with tablet PCs on which the honesty task was presented to the subjects in the following way: The two lotteries were shown as two urns, each containing three balls as shown in the top right panel of Figure 1. Two of the balls in each urn always had the same value, whereas the third ball always had a different, higher value. The urn pairs were chosen such that one urn always had a larger spread between the high and the low outcome, but the two urns always yielded the same expected value. The full list of all five lottery pairs is shown in Table 1.[8]

In order to physically draw from their chosen digital urn, participants were given a small cotton pouch containing three wooden tiles with different symbols on them before the experiment (a circle, a star, and a diamond). After choosing an urn, participants read that the computer had now randomly matched each of

---

[7] The whole set of questions is shown in Appendix A along with further details on all other procedures. These questions are partially based on the short dark triad questionnaire by Jones & Paulhus (2014) and the question on self-perceived morality from Mazar et al. (2008). More elaborate measures were not feasible because of a very strict time constraint: Many people who want to go to a bar are easily put off by the prospect of an experiment that lasts more than just a few minutes. We therefore had to restrict the experiment to its core features such that it would last no longer than 10 minutes (when recruiting, we announced an average duration of 6 to 8 minutes).

[8] Also note that, under risk aversion, an honest subject should never choose the risky urn due to the equality of expected values. Choosing the risky urn could thus already be indicative of the intention to lie in the next step, yet, our data do not support this hypothesis.

the three balls in the chosen urn with one of the three tiles in their pouch. They were then instructed to draw a tile from the pouch and press continue to see the matching of balls to tiles. On the next screen they saw the matching and were instructed to specify their income according to the tile they had just drawn by tapping any of the three blue rectangles (see the two bottom panels of Figure 1). At this step, participants could lie and choose a different payoff than the one they had drawn. One of the five rounds was randomly chosen to be relevant for payment.

There is an increasingly applied alternative method in which participants draw a tile *mentally* instead of physically before they observe the matching of tiles to payoffs (see, e.g., Shalvi, 2012; Jiang, 2013; Shalvi & De Dreu, 2014; Barfort et al., 2015; Kajackaite & Gneezy, 2015; and Potters & Stoop, 2016, for other studies using such mind games). This method comes with the advantage that there can be no doubt for subjects that only they can know which tile they have drawn and that they are not secretly monitored. However, it also requires subjects to remember their mentally drawn tile. This could be critical in our setting: More intoxicated subjects might actually be worse at remembering or could more easily justify a lie to themselves because they can more reasonably claim to have forgotten their mental draw. This can be ruled out in our design as participants still hold the drawn tile in their hand as they enter their report.

161 subjects (55% male, average age 22.7 years) were recruited in front of two bars in Konstanz and Ulm. Both treatments were run in both locations on two consecutive days. 59.6% of the subjects participated in the *intoxicated* condition. The main part of the experiment inside the van lasted about 8 minutes and average earnings were about 8.50 Euros.

TABLE 1—PAYOFFS OF THE FIVE LOTTERY PAIRS (IN EXPERIMENTAL CURRENCY UNITS).

Safe Urn		Risky Urn	
π_{High}	π_{Low}	π_{High}	π_{Low}
40	37	50	32
42	39	52	34
44	41	54	36
46	43	56	38
48	45	58	40

Notes: In all cases, the higher payoff, π_{High}, always had a probability of 1/3 whereas the lower payoff, π_{Low}, always had a probability of 2/3. The exchange rate of ECU to Euros was 10 to 1.

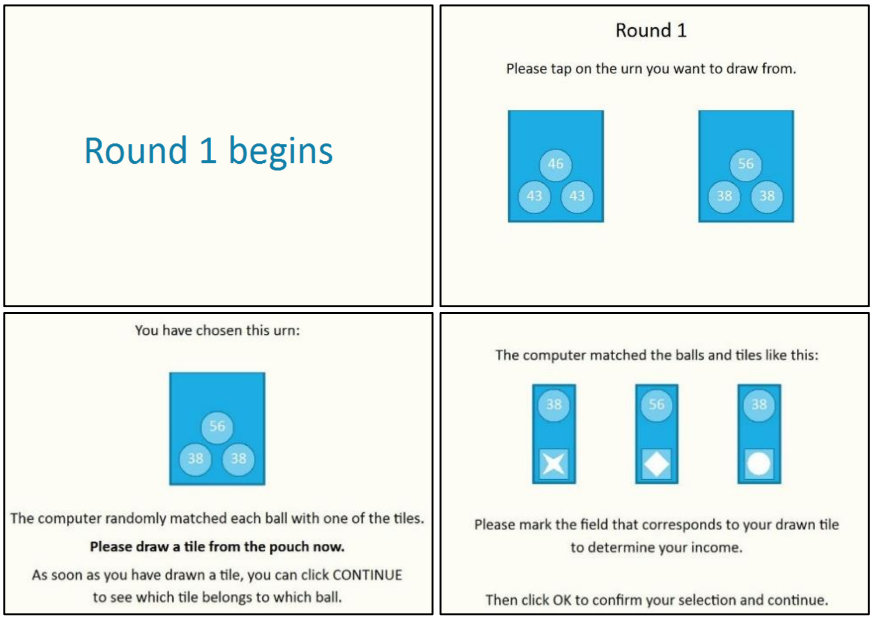


FIGURE 1. SEQUENCE OF SCREENS IN EACH ROUND OF THE HONESTY TASK.

Notes: Translated from German original. For the sake of space, the "continue" button in the 3rd screen and the "ok" button in the 4th screen are cut out here. They were both positioned in the bottom right corner of the respective screens. The sequence of the screens was top left (800 *ms*), top right, bottom left, bottom right. The high payoff was always displayed as the top ball in the urns, but its position on the 4th screen was randomly determined in every round (the same holds for the positions of the three tile-symbols on the 4th screen).

IV. Results

A. *Main Analysis*

Figure 2 shows the cumulative distributions of the measured breath alcohol contents (BAC) in the two conditions before participants entered the bar (BAC1) and after they left (BAC2). The two populations did not differ with respect to their alcohol consumption: The CDFs of the first measurements in the two conditions are almost identical and the same holds for the second measurements (the respective Kolmogorov-Smirnov tests are far from significant, both

p-values > .65). There is a clear difference between the two relevant BAC-measurements though: With the exception of just one observation, the CDF of the first measurement in the *sober* condition (solid blue) is always to the left of the CDF of the second measurement in the *intoxicated* condition (solid red). Obviously, the respective Kolmogorov-Smirnov test is also highly significant ($p < 0.001$) and we can conclude that our treatment condition presents a valid instrument for alcohol intake.

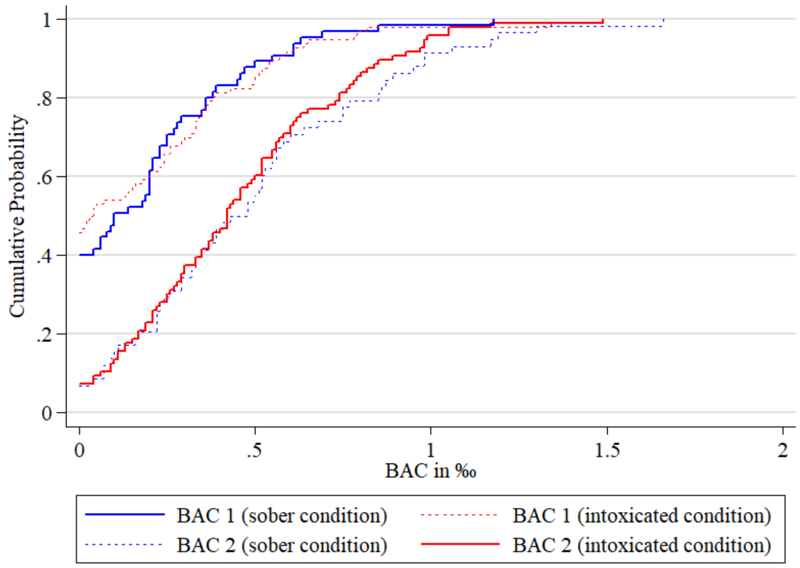


FIGURE 2. CUMULATIVE DISTRIBUTIONS OF MEASURED BREATH ALCOHOL CONTENTS IN ‰.

Notes: The blue (red) lines show the BAC measurements of the participants in the *sober* (*intoxicated*) condition who participated in the experiment before entering (after leaving) the bar. The lines are solid when they show the relevant BAC given the condition, i.e., the measurement which was taken in direct succession with the experiment.

For a first impression of the effect of acute alcohol intoxication on honesty, we now compare the distribution of the number of reported high outcomes in the two treatment conditions. Figure 3 shows the respective CDFs as well as the expected distribution under the theoretic benchmark of complete honesty. People lied more in the *sober* condition than in the *intoxicated* condition. In fact, the *sober* CDF even first order stochastically dominates the *intoxicated* CDF. This visual result is also corroborated by a robust rank order test (Fligner & Policello, 1981) comparing the distributions of the number of high reports in the two treatments (equality of distributions is rejected with $p = 0.051$). When we compare the observed distributions with the theoretical benchmark of full honesty using one-sample Kolmogorov-Smirnov tests, we find that only the *sober* CDF is significantly different ($p < 0.05$), whereas there is no significant evidence for dishonesty in the *intoxicated* condition ($p > 0.2$). As can be seen in Figure 6 in Appendix B, the treatment differences are particularly strong in the range of significantly many high reports: The share of participants that reported four or more high outcomes in the *sober* condition (20%) is more than three times as high than in the *intoxicated* condition (6.2%).[9]

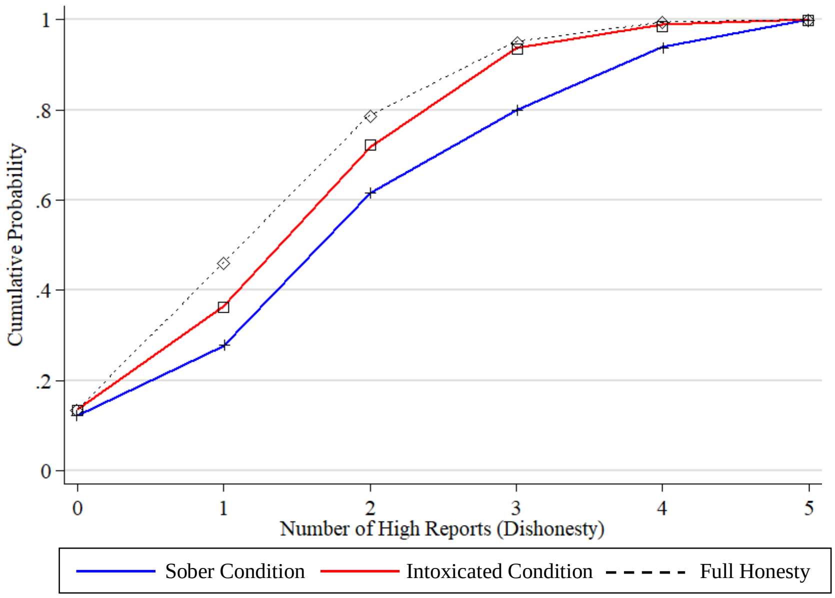


FIGURE 3. CUMULATIVE DISTRIBUTION OF THE NUMBER OF HIGH REPORTS IN THE TWO CONDITIONS.
Notes: The dashed line shows the theoretical benchmark of complete honesty (binomial distribution with 5 trials and a success probability of 1/3).

We now move to a parametric analysis of our data and regress the decision whether to report a high outcome on the level of intoxication as measured with the breathalyzer. As discussed in the introduction, the BAC measurements should be highly correlated with our participants' general drinking habits. We can make use of this when we want to explore differences in honesty preferences

---

[9] The probability of actually drawing the high outcome at least four times is equal to $p = \sum_{k=4}^{5} \binom{5}{k} \frac{1}{3}^k \frac{2}{3}^{5-k} = \frac{10}{243} + \frac{1}{243} \approx 0.045$, i.e., a subject that reported four high outcomes is a significant liar at the 5% level. Similarly, a subject that reported five high outcomes is a significant liar at the 1% level (the probability of five high outcomes is $\frac{1}{243} \approx 0.004$).

based on differences in drinking habits. Yet, we also have to control for it by using our treatment conditions as an instrumental variable when we want to explore the effect of acute intoxication on dishonesty.

Model 1 in Table 2 shows the results of a linear probability model in which we regressed the decision to report a high outcome on the second BAC measurement and some control variables (age, gender, group size, location, and money carried in wallet). We find a significantly positive effect. The left panel of Figure 4 depicts a binned scatter plot of the same model.[10] The second BAC measurement, BAC2, is also our best proxy for general drinking habits because there is more variation than in BAC1: More than 40% of participants arrive completely sober but only 7% leave with a BAC of zero. We nevertheless ran the same regression using BAC1 as a robustness check and again find a significantly positive effect (see model 2 in Table 2).

As pointed out above, we have to use an instrumental variable approach in order to correctly estimate the effect of acute intoxication. We use a standard two-stage least-squares IV regression (2SLS-IV) with the treatment condition as the instrument for the relevant BAC. The results are shown numerically in model 4 in Table 2 and graphically in the right panel of Figure 4. Here, we again find the significantly negative effect we already know from the non-parametric analysis: People lie less when they are more intoxicated during the experiment.

Model 3 in Table 2 shows what our results would be if we simply conducted the experiment without the two treatments but with participants under (endogenously) varying levels of intoxication at the time of the experiment. It shows a positive but insignificant ($p > .2$) effect of the relevant BAC (the one participants had *during* the experiment) on the probability of reporting a high outcome. Ignoring the correlation between acute intoxication and general drinking habits thus would have led us to underestimate the effect of general drinking habits

[10] Binning is necessary for visual clarity since the outcome variable *ReportedHigh* is binary.

and completely neglect the true (negative) effect of acute intoxication on dishonesty.

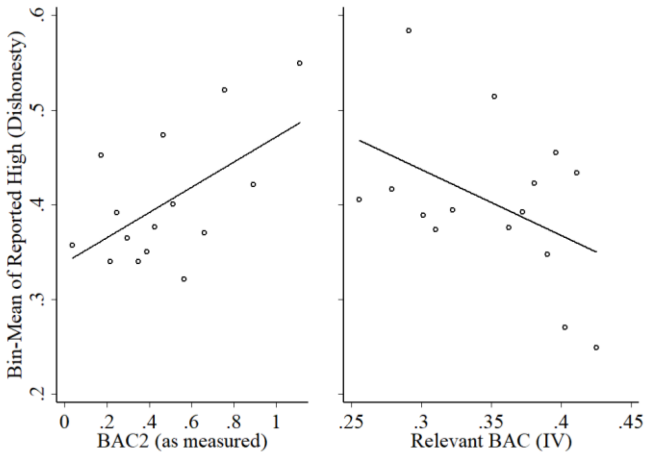


FIGURE 4. BINNED SCATTER PLOTS OF THE MEAN OF THE BINARY OUTCOME VARIABLE "REPORTED HIGH" ON MEASURED BAC (LEFT) AND BAC INSTRUMENTED BY TREATMENT (RIGHT).

Notes: The left graph shows the effect of the second BAC measurement on the likelihood of a high report as predicted by a linear probability model in which we additionally controlled for some nuisance variables (age, gender, group size, location, money carried in wallet). To construct the binned scatter points, the data was first divided into 15 bins containing equally many observations. After controlling for the nuisance variables, the means of BAC2 and of the binary outcome variable *ReportedHigh* are computed for each bin and are then displayed as one point in the graph. The linear fit corresponds to the slope coefficient of the measured BAC2 in the same model (see model 1 in Table 2).
The right graph is constructed in the same way but using a 2-stage least-squares model (with the same control variables) in which the treatment condition was used as an instrumental variable for the *relevant* BAC. The *x*-axis of the right graph thus reflects the results of the first stage regression (controlled for the mentioned nuisance variables), i.e., the predicted values of the relevant BAC—the different scaling on the *x*-axis is thus due to the natural loss of precision in IV models. The results of all first stage regressions are shown in Table 5 in Appendix B. The results of the underlying 2nd stage model are shown in model 4 in Table 2. For a more encompassing explanation on the construction of binned scatter plots, see Stepner (2014).

TABLE 2—REPORTED HIGH OUTCOME REGRESSED ON BAC.

Dep. Variable:	Reported High Outcome (Dishonesty, Binary)				
Model:	(LPM)			(2SLS-IV)	
	(1)	(2)	(3)	(4)	(5)
BAC2	.133**				
	(.061)				
BAC1		.149**			
		(.063)			
Relevant BAC			.101	-.694*	-.309**
			(.081)	(.370)	(.162)
Constant	.362**	.370**	.381***	.530***	.511***
	(.180)	(.177)	(.176)	(.187)	(.063)
Controls	Yes	Yes	Yes	Yes	No
N	805	805	805	805	805

Notes: The number of observations corresponds to the five rounds of all 161 subjects. Robust standard errors clustered at the subject level in parentheses. Controls: Gender, age, money carried in wallet, group size, location. Relevant BAC is instrumented by the treatment condition in models (4) and (5).
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

As a robustness check, we repeated our analyses using probit instead of linear probability models and the two-step IV probit control function approach introduced by Rivers & Vuong (1988) instead of the 2SLS-IV models. Because the results are virtually identical, we do not report them here but in Appendix C in which we also report the analyses in which we included the time of participation as a control for people's tiredness, which did not change our results either.

B. Further Results

We now separately analyze the behavior of all those participants who stayed sober in the bar, i.e., those who had a BAC of 0.00 in the second measurement. These participants were exposed to the same aspects of the bar visit, except that they did not consume any alcohol. Doing this can thus help us disentangle whether the found effect is mainly driven by the social aspects or by the amount of alcohol consumed. The mean number of high reports is indeed slightly lower after the bar visit also for participants who stayed sober (2.25 high reports in the *sober* condition versus 2 in the *intoxicated* condition), however, the corresponding robust rank order test is far from being rejected at conventional levels of significance ($p > 0.45$).[11] Since there is no variation in the relevant BAC in the restricted sample, the first stage regressions of the respective instrumental variable approaches become redundant. The correct regression models thus reduce to regular Probit and OLS models with a dummy variable for the treatment condition instead of the (instrumented) relevant BAC. These models yield similar results: Spending time in a bar with friends is estimated to slightly reduce dishonesty but not significantly so (see Table 6 in Appendix B). Suggesting that a night at the bar per se is not sufficient to increase honesty but that people indeed have to consume alcohol for this effect to emerge.

---

[11] This could of course be due to the small number of observations in this restricted sample (only about 7% of participants stayed completely sober in each condition). However, the result remains far from significant even if we relax the restriction regarding sobriety and include, for instance, all participants with $BAC2 \leq 0.1$.

We ran additional regressions to test whether the effect of acute alcohol intoxication is actually driven or moderated by differences in subjects' cognitive ability or morality. Cognitive abilities are widely known to suffer from acute intoxication. Results as to how intelligence relates to honesty are mixed but studies in which cognitive load was varied find increased honesty under higher load (see the literature reviewed in section II). Since alcohol intoxication can be seen as a form of increased load, our result that alcohol decreases dishonesty is in line with these findings.

Model (1) in Table 3 includes our measure for cognitive ability and model (2) additionally includes its interaction term with the instrumented relevant BAC. Even though we do find a significantly negative effect of cognitive ability on dishonesty, also the coefficients on the relevant BAC stay significantly negative. If anything, the effect of acute intoxication even becomes more pronounced compared to its counterpart where we did not control for cognitive ability (model (4) in Table 2.) We find no evidence for a moderating effect of cognitive ability as the coefficients on the interaction term does not reach significance.

As discussed in the literature review, alcohol intake has also been found to lower moral standards, and inattention to moral standards has been found to negatively affect honesty. Taken together, these findings would suggest a positive overall effect of alcohol intoxication on dishonesty. This is not in line with what we find. We included our measure for morality in model (3) and additionally its interaction term with the instrumented relevant BAC in model (4) in Table 3. The respective coefficients are small and never significantly different from zero in either model. The negative effect of the instrumented relevant BAC stays significant and similar in size compared to the model without controls for morality in Table 2. We thus find no evidence that our main effect is driven or moderated by lowered moral standards.[12]

---

[12] We conducted the same robustness checks for these models as for those in the main analysis (i.e., estimating 2SIV-Probit models instead of the linear probability models reported here). Because the results do not change, we report them in Table 9 in Appendix C.

One might expect that some subjects would try to be dishonest only when it is particularly worth it but not when the relative gains of being dishonest are comparatively small. To test this, we also included the size of the high payoff of the chosen urn as an explanatory variable in our models. However, we found no evidence in support of such strategically selective dishonesty. A potential reason for this is that subjects did not know the games in advance and played them in individually randomized orders. This means that, at least in the first rounds, they could not know whether the potential gains of dishonesty from the current lottery were relatively high or low. Further, our lotteries were chosen such that the *absolute* difference between the high and the low outcome was constant across all lotteries, making our design less suitable to study effects of stake size.

TABLE 3—REPORTED HIGH OUTCOME REGRESSED ON BAC, COGNITIVE ABILITY, AND MORALITY.

Dep. Variable:	Reported High Outcome (Dishonesty, Binary)			
Model:	(2SLS-IV)			
	(1)	(2)	(3)	(4)
Relevant BAC	-.703*	-1.091**	-.669*	-.631*
	(.373)	(.451)	(.372)	(.367)
Cog. Ability	-.028*	-.060**		
	(.015)	(.026)		
BAC×Cog.Ab.		.087		
		(.055)		
Morality			-.002	.001
			(.003)	(.007)
BAC×Moral.				-.009
				(.015)
Constant	.676***	.827***	.542***	.523***
	(.212)	(.230)	(.187)	(.185)
Controls	Yes	Yes	Yes	Yes
N	805	805	805	805

Notes: Robust standard errors clustered at the subject level in parentheses. Controls: Gender, age, money carried in wallet, group size, location. Relevant BAC instrumented by treatment condition in all models.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Recall that, in the honesty game, the two urns had the same expected value in each of the five rounds. This directly implies that a risk-averse subject should never choose the risky urn when she intends to be honest in the next step. In comparison, in the coin-flip task, a risk-neutral player should always invest everything whereas the optimal bet size decreases with increasing risk-aversion.

More than 90% of our subjects displayed risk-aversion in the coin-flip task and did not invest their whole endowment, yet, only 12% never chose the risky urn (on average, the risky urn was chosen in 46% of all rounds). To test the hypothesis that people who choose the risky urn would be more likely to report a high outcome, we included a dummy variable indicating the choice of the risky urn in our models. The effect of choosing the risky urn on the probability of reporting a high outcome was estimated to be very small and far from significant in all models. Also all non-parametric tests we ran to compare the distributions of the number of high reports of those who (almost) always chose the risky urn with those who chose it less often were all insignificant. Choosing the risky urn thus cannot reliably indicate the intention to lie in our sample. Finally, we investigated whether risk-preferences as measured by the coin-flip task correlated with behavior in our honesty game. Note first, that we could not detect any significant relationship between BAC and the bet size in the coin-flip task. This is in line with the findings reported by Breslin et al. (1999) and the generally inconclusive results in the literature on the effects of alcohol on risk preferences. We still tested whether behavior in the coin-flip task could explain urn choices or high reports, but also here, the estimated coefficients were always small and never significant.

V. Conclusion

We have investigated the influence of social drinking on people's honesty with a lab-in-the-field experiment conducted in front of bars. We have two main findings: People become more honest, the more they had to drink, yet, people who drink more in general, are *less* honest in general. Despite the field setting, we can make a causal claim because we exogenously imposed two treatments which only differed with respect to the timing of the experiment. Our design thus also entails a methodological contribution as it can easily be extended to cleanly study the effects of social drinking on other aspects of economic decision making. We can control for the potential endogeneity problem that other field studies typically entail (acute intoxication is correlated with general drinking habits) by using the treatment condition as an instrumental variable for the measured breath alcohol content.

We can further show that the effect cannot be solely attributed to the non-alcohol related aspects of the bar visit as participants who stayed sober did not become significantly more honest after their stay inside the bar. Previous literature indicated potential moderating effects of cognitive ability and moral standards as they have both been linked to both alcohol intoxication and honesty preferences. We find a small positive effect of cognitive ability on honesty but no evidence for any moderating effects.

Our study focused on dishonesty for gaining a personal economic advantage. There are of course many other dimensions of dishonesty which are not captured well by such a design and that could be the focus of future research. For instance, it seems plausible that people's willingness to exaggerate a story to make it more interesting or fun could increase under the influence of alcohol. This kind of dishonesty is not measured here. Similarly, changes in lying behavior after alcohol consumption could also be dependent on the effects of one's dishonesty on others. For example, white lies as studied by Erat & Gneezy (2012), could become *more* frequent when people are intoxicated than when sober since alcohol is known to enhance positive affect and social bonding (see, e.g., the review by Sayette, 2017). The stronger emotional bond experienced by social drinkers could thus lead to more dishonesty when lying is beneficial to the liar's peers.

Appendix

A. Procedural Details

We set up our mobile lab inside a van with separator screens between seats such that up to six subjects could participate at a time. The van was parked in front of a bar and people who came to enter the bar were invited to participate in a brief experiment on the effects of alcohol on risky decisions. Those that agreed to participate immediately received a show-up fee of 2 Euros and a ticket with a three-digit ID number which they needed to collect their earnings when they left the bar. We then measured their level of alcohol intoxication with a breathalyzer.

The next step was the main part of the experiment: Each subject received a tablet computer and a cotton pouch containing three wooden tiles which were later used for the honesty task. They took a seat in the van and first entered their ID number and then went through the following stages: 1. A consent form in which participants agree to the scientific use of their data, 2. A single item risk elicitation, 3. Five rounds of the honesty game, 4. A brief test of their cognitive ability, 5. Five 11-point Likert-scaled questions on morality. Finally, there was a brief socio-economic questionnaire. While the games in stages 2 and 3 were incentivized, participants received a flat payment of 1 Euro for stages 4 and 5 and the final questionnaire. Once they were done, we measured their breath alcohol content for a second time after which they received their earnings from the experiment. Participants knew from the beginning that they will never learn their measured level of alcohol intoxication—neither after the first nor the second measurement. This was done to avoid a potential influence of the knowledge about the later test on drinking behavior.

There were two treatment conditions which differed only with respect to after which stage of the experiment participants entered the bar. In the *intoxicated* condition, participants went inside the bar right after they received the show-up fee, their ID number, and had completed the first alcohol measurement (BAC1).

These subjects completed the rest of the experiment right after they left the bar. In the *sober* condition, the participants went inside the bar after the main part of the experiment and completed the socio-economic questionnaire and the second alcohol measurement (BAC2) after they came back out. Figure 5 shows an overview of the course of the experiment for both treatments. The two treatments were run on different days such that inside the bar either all (*sober* treatment) or none of the participants (*intoxicated* treatment) already knew the honesty task.

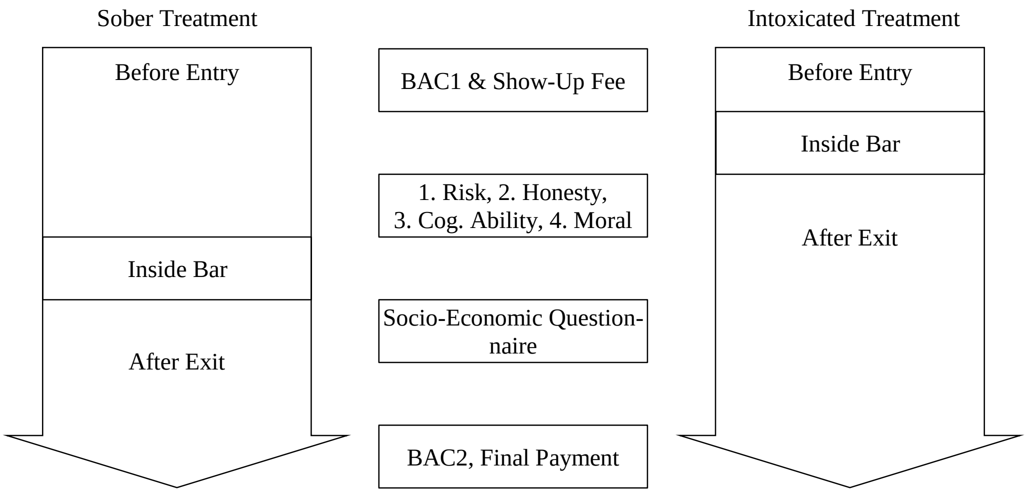


FIGURE 5. COURSE OF THE EXPERIMENT, SOBER TREATMENT ON THE LEFT, INTOXICATED ON THE RIGHT.

Our single item risk elicitation task was similar to that used by Gneezy & Potters (1997): Subjects received 10 points (worth 1 Euro) to bet on a coin flip. For the points they bet, they had a 50-50 chance to either get back nothing or 2.5 times the amount bet. All points they did not bet, they kept for sure. The coin flip was conducted by the experimenter directly before participants received their earnings.

Our test for participants' cognitive ability consisted of ten mathematical equations for which participants had to indicate whether they were true or false. They had 5 seconds for each equation and not answering in time was counted as incorrect. The full set of equations is shown in Table 4. Our measure for cognitive ability simply corresponds to the number of correctly solved equations.

TABLE 4—EQUATIONS USED FOR THE COGNITIVE ABILITY TEST.

Equation No.	Equation	Equation No.	Equation
1	989 – 789 = 200 (true)	6	572 - 389 = 183 (true)
2	247 – 32 = 216 (false)	7	361 - 194 = 177 (false)
3	774 – 561 = 213 (true)	8	386 - 256 = 140 (false)
4	856 – 638 = 218 (true)	9	753 - 379 = 374 (true)
5	295 – 160 = 145 (false)	10	479 - 135 = 324 (false)

Notes: All subjects saw the equations in the above order which was randomly chosen before the experiment. Subjects correctly solved between 0 and 9 equations. The average number of correct answers across treatments is 4.3 with a standard deviation of 1.8.

Finally, our measure of morality comprised the following five 11-point Likert-scaled questions: (1) *I would stop at nothing to reach my goals.* (2) *You observe someone losing a 5€ bill a few meters in front of you. It is morally okay to take the bill and keep it.* (3) *There are many situations in which legal rules can be stretched a little.* (4) *Being moral is more important than acting for your own personal interest.* (5) *Compared to yesterday: How moral do you feel today?* Questions (1) through (4) ranged from "Do not agree at all" (-5) to "Completely agree" (+5) and question (5) ranged from "Much less moral" (-5) to "Much more moral" (+5). Answers to the first three questions negatively enter our measure for morality and answers to questions (4) and (5) enter positively.

B. Further Analyses

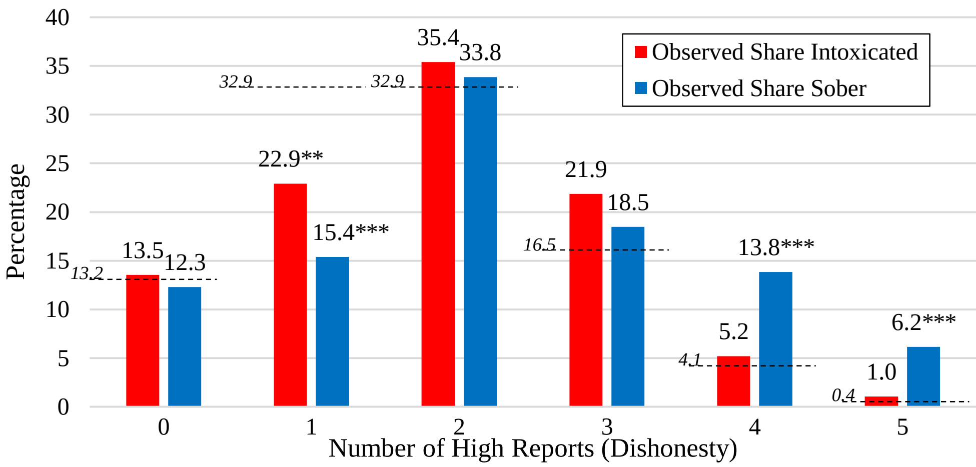


FIGURE 6. PERCENTAGE OF NUMBER OF HIGH REPORTS IN THE TWO TREATMENT CONDITIONS.

Notes: The dashed lines show the expected shares of the respective number of high reports under complete honesty. Asterisks indicate the significance levels of two-sided binomial tests that the observed percentages differ from the expected percentages under complete honesty (*** $p < 1\%$, ** $p < 5\%$).

TABLE 5—FIRST STAGE REGRESSIONS OF ALL 2SLS-IV MODELS.

Dep. Variable:	Relevant BAC			
Model:	(2SLS-IV, 1st Stage)			
	(1)	(2)	(3)	(4)
Intoxicated Cond.	.217***	.254***	.216***	.219***
	(.047)	(.044)	(.046)	(.047)
Male	.173***		.187***	.169***
	(.044)		(.045)	(.047)
Age	-.005		-.007	-.004
	(.005)		(.005)	(.005)
Money in Wallet	-.0002***		-.0002***	-.0002***
	(.000)		(.000)	(.000)
Group Size	.006**		.007***	.006**
	(.002)		(.002)	(.002)
Location	.068		.070*	.066
	(.043)		(.042)	(.044)
Cog. Ability			-.030**	
			(.013)	
Morality				-.002
				(.004)
Constant	.193	.198***	.359**	.203
	(.139)	(.031)	(.159)	(.146)
N	161	161	161	161

Notes: Robust standard errors in parentheses. One observation per subject. *Intoxicated Cond.* is a dummy variable that is equal to one for all subjects in the *intoxicated* condition and zero otherwise. Location is equal to one when the experiment was conducted in Ulm and zero when it was conducted in Konstanz. Models 1 and 2 correspond to the first stage regressions of models 2 and 3 in Table 2, respectively. Likewise, models 3 and 4 correspond to those of models 1 and 2 in Table 3.
***Significant at the 1 percent level.
**Significant at the 5 percent level.

TABLE 6—REPORTED HIGH OUTCOME REGRESSED ON TREATMENT DUMMY (SOBER PARTICIPANTS ONLY).

Dep. Variable:	Reported High Outcome (Dishonesty, Binary)			
Model:	(LPM)		(Probit)	
	(1)	(2)	(3)	(4)
Intoxicated Cond.	-.050	-.056	-.128	-.160
	(.150)	(.095)	(.378)	(.249)
Constant	.450***	.184	-.126	-.737
	(.138)	(.610)	(.344)	(1.569)
Controls	No	Yes	No	Yes
N	55	55	55	55

Notes: Robust standard errors clustered at the subject level in parentheses. Data set restricted to all those participants who stayed sober in the bar (BAC2 = 0). The control variables are gender, age, money carried in wallet, group size, and location.
***Significant at the 1 percent level.

C. Robustness Checks

Linear probability models as reported in our main analysis generally run the risk of resulting in inconsistent and unbiased estimators (see, e.g., the textbook on linear probability, logit, and probit models by Aldrich & Nelson, 1984, for a detailed discussion). Horrace & Oaxaca (2006) formally derive conditions that have to hold for consistency and unbiasedness of LPM estimates. In fact, these conditions hold for each model we estimated with our sample. But since this could be due to our sample being restricted, we nevertheless ran probit regressions as robustness checks for our LPMs and used the two-step IV probit (2SIV-Probit) control function approach introduced by Rivers & Vuong (1988) as a robustness check for our 2SLS-IV models. We explain this procedure in a little more detail below. The results of these models are reported in Table 7 which mirrors Table 2 from the main analysis, i.e., model 1 here is the robustness check for model 1 there and so forth.

The results from the standard linear probability models are very robust. The estimated average partial effects of the corresponding probit models are almost exactly the same and never differ before the third decimal place. Before we discuss the results of the two-step IV-Probit models, we will give a brief explanation of how they work: As in the standard 2SLS-IV case, the first stage regression in the case of the 2SIV-Probit model consists of an OLS regression of the variable to be instrumented on the instrumental variable and all other exogenous variables. Instead of using the fitted values of the first stage as a regressor in the second stage probit model, the endogenous variable remains in the model and the residuals from the first stage are added as an explanatory. This procedure not only yields consistent estimates up to a (known) common scale but has the main advantage of a built-in test for the exogeneity of the (potentially) endogenous variable.[13] This test simply corresponds to the standard probit *z*-statistic on the first stage residuals as a regressor in the second stage. The bottom two rows of columns (4) and (5) in Table 7 show that the exogeneity of the relevant BAC is clearly rejected. This lends additional strong support to the necessity of the instrumental variable approach. We further see that also the effects found with the simpler 2SLS-IV approach from the main analysis are robust: The estimated

[13] See Wooldridge (2010, pp. 585-594) for a detailed explanation of the procedure.

average partial effects of the (instrumented) relevant BAC on the likelihood of a high report are significantly negative and very similar in size in the 2SIV-Probit models as compared to the results in the main analysis.

TABLE 7—REPORTED HIGH OUTCOME REGRESSED ON BAC (AVERAGE PARTIAL EFFECTS).

Dep. Variable:	Reported High Outcome (Dishonesty, Binary)				
Model:	(Probit, APEs) (1)	(2)	(3)	(2SIV-Probit, APEs) (4)	(5)
BAC2	.131**				
	(.060)				
BAC1		.148**			
		(.061)			
Relevant BAC			.101	-.608*	-.296**
			(.080)	(.350)	(.143)
Controls	Yes	Yes	Yes	Yes	No
N	805	805	805	805	805
Exogeneity test (coefficient on 1st stage residuals in 2nd stage reg.)				.952**	2.168**
				(.419)	(.936)

Notes: Standard errors for underlying models robust and clustered at the subject level. To obtain standard errors for the average partial effects that we report here, the Delta-Method was used in case of models (1) – (3). For models (4) and (5) we obtained standard errors for the average partial effects via bootstrapping the procedure (1000 replications). The results of the 2SIV-Probit's built-in test for exogeneity of the instrumented variable is shown in the two bottom rows. Controls: Gender, age, money carried in wallet, group size, location. Relevant BAC is instrumented by the treatment condition in models (4) and (5).
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 8 below reports the same models as those reported in Table 2 in the main analysis except that they additionally include the time of participation to control for participants' tiredness. In all five specifications the effect of the time of participation (measured in minutes since noon) is estimated to be very small and it is never significant. The effect of the measured BAC in models (1) – (3) remains virtually unchanged compared to the main analysis. Similarly, the effect of the relevant BAC instrumented by the treatment condition in models (4) and (5) remain significantly negative (but are even larger in size). We thus conclude that participants' tiredness does not drive our results.

TABLE 8—REPORTED HIGH OUTCOME REGRESSED ON BAC AND TIME OF PARTICIPATION.

Dep. Variable:	Reported High Outcome (Dishonesty, Binary)				
Model:	(LPM)			(2SLS-IV)	
	(1)	(2)	(3)	(4)	(5)
Participation Time	-.000*	-.000	-.001	.001	.001
	(.000)	(.000)	(.000)	(.001)	(.001)
BAC2	.144**				
	(.061)				
BAC1		.149**			
		(.061)			
Relevant BAC			.148*	-1.352*	-.637**
			(.085)	(.808)	(.315)
Constant	.441**	.440*	.471**	.467***	.462***
	(.195)	(.190)	(.190)	(.138)	(.043)
Controls	Yes	Yes	Yes	Yes	No
N	805	805	805	805	805

Notes: Robust standard errors clustered at the subject level in parentheses. Controls: Gender, age, money carried in wallet, group size, location. Relevant BAC is instrumented by the treatment condition in models (4) and (5).
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 9 below shows the results of the 2SIV-Probit models that correspond to the 2SLS-IV models in Table 3 in the further analyses section. Note that the partial effects of interaction terms were corrected for the immanent non-linearity of these models as called for by Ai & Norton (2003). For usual probit models, this can be done using the software package "inteff" (Norton, Wang & Ai, 2004). But since the coefficient estimates of the 2SIV-Probit models are only consistent up to a common scale, one first has to compute the unscaled coefficients after the probit estimation and then use these to compute partial effects. We applied the correction for the non-linearity to the unscaled coefficients by adapting the code by Norton et al. (2004) accordingly and obtained standard errors via bootstrapping as before.

TABLE 9—REPORTED HIGH OUTCOME REGRESSED ON BAC, COGNITIVE ABILITY, AND MORALITY (AVERAGE PARTIAL EFFECTS).

Dep. Variable:	Reported High Outcome (Dishonesty, Binary)			
Model:	(2SIV-Probit, APEs)			
	(1)	(2)	(3)	(4)
Relevant BAC	-.619**	-.769**	-.589*	-.594*
	(.313)	(.342)	(.331)	(.347)
Cog. Ability	-.025*	-.037**		
	(.014)	(.018)		
BAC×Cog.Ab.		.041		
		(.031)		
Morality			-.002	-.003
			(.004)	(.005)
BAC×Moral.				.002
				(.008)
Controls	Yes	Yes	Yes	Yes
N	805	805	805	805
Exogeneity test (coefficient on 1st stage residuals in 2nd stage reg.)	2.174**	2.298**	2.101**	2.084**
	(.947)	(.951)	(.937)	(.938)

Notes: Standard errors for underlying models robust and clustered at the subject level. Standard errors for the average partial effects that we report here, obtained via bootstrapping the procedure (1000 replications). The results of the built-in test for exogeneity of the instrumented variable is shown in the two bottom rows. Controls: Gender, age, money carried in wallet, group size, location. Relevant BAC is instrumented by the treatment condition in all models.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

References

Abeler, J., Nosenzo, D., & Raymond, C. (2016). Preferences for truth-telling. *CESifo Working Paper Series, No. 6087*, 1-118.

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters, 80*(1), 123-129.

Aldrich, J. H., & Nelson, F. D. (1984). *Linear Probability, Logit, and Probit Models* (Vol. 45). Newbury Park, London, New Delhi: Sage.

Anderson, L. R., & Mellor, J. M. (2008). Predicting health behaviors with an experimental measure of risk preference. *Journal of Health Economics, 27*(5), 1260-1274.

Arbel, Y., Bar-El, R., Siniver, E., & Tobol, Y. (2014). Roll a die and tell a lie– What affects honesty? *Journal of Economic Behavior & Organization, 107*, 153-172.

Arthur, W., & Day, D. V. (1994). Development of a short form for the Raven Advanced Progressive Matrices Test. *Educational and Psychological Measurement, 54*(2), 394-403.

Barfort, S., Harmon, N., Hjorth, F., & Olsen, A. L. (2015). Dishonesty and selection into public service in Denmark: Who runs the world's least corrupt public sector? *University of Copenhagen Discussion Paper Series, 15*(12), 1-56.

Barsky, R. B., Juster, F. T., Kimball, M. S., & Shapiro, M. D. (1997). Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. *Quarterly Journal of Economics, 112*(2), 537-579.

Bègue, L., Bushman, B. J., Zerhouni, O., Subra, B., & Ourabah, M. (2013). 'Beauty is in the eye of the beer holder': People who think they are drunk also think they are attractive. *British Journal of Psychology*, *104*(2), 225-234.

Biziou-van-Pol, L., Haenen, J., Novaro, A., Liberman, A. O., & Capraro, V. (2015). Does telling white lies signal pro-social preferences? *Judgment and Decision Making, 10*(6), 538-548.

Bregu, K., Deck, C., Ham, L., & Jahedi, S. (2017). The effects of alcohol use on economic decision making. *Southern Economic Journal, 83*(4), 886-902.

Breslin, F. C., Sobell, M. B., Cappell, H., Vakili, S., & Poulos, C. X. (1999). The effects of alcohol, gender, and sensation seeking on the gambling choices of social drinkers. *Psychology of Addictive Behaviors*, *13*(3), 243-252.

Burghart, D. R., Glimcher, P. W., & Lazzaro, S. C. (2013). An expected utility maximizer walks into a bar. *Journal of Risk and Uncertainty, 46*(3), 215-246.

Cappelen, A. W., Sørensen, E. Ø., & Tungodden, B. (2013). When do we lie? *Journal of Economic Behavior & Organization, 93*, 258-265.

Capraro, V. (2017a). Does the truth come naturally? Time pressure increases honesty in one-shot deception games. *Economics Letters, 158*, 54-57.

Capraro, V. (2017b). Who lies? A meta-analysis of the effect of sex, age, and education on honesty. *Mimeo*.

Charness, G., & Dufwenberg, M. (2006). Promises and partnership. *Econometrica, 74*(6), 1579-1601.

Childs, J. (2012). Gender differences in lying. *Economics Letters, 114*(2), 147-149.

Cohen, T. R., Gunia, B. C., Kim-Jun, S. Y., & Murnighan, J. K. (2009). Do groups lie more than individuals? Honesty and deception as a function of strategic self-interest. *Journal of Experimental Social Psychology, 45*(6), 1321-1324.

Cohn, A., & Maréchal, M. A. (2017). Laboratory measure of cheating predicts school misconduct. *Mimeo*.

Cohn, A., Maréchal, M. A., & Noll, T. (2015). Bad boys: How criminal identity salience affects rule violation. *Review of Economic Studies, 82*(4), 1289-1308.

Conrads, J., Irlenbusch, B., Rilke, R. M., & Walkowitz, G. (2013). Lying and team incentives. *Journal of Economic Psychology, 34*, 1-7.

Corazzini, L., Filippin, A., & Vanin, P. (2014). Economic behavior under alcohol influence: An experiment on time, risk, and social preferences. *IZA Discussion Paper Series, No. 8170*.

Curtin, J. J., & Fairchild, B. A. (2003). Alcohol and cognitive control: Implications for regulation of behavior during response conflict. *Journal of Abnormal Psychology, 112*(3), 424-436.

Dai, Z., Galeotti, F., & Villeval, M. C. (2017). Cheating in the lab predicts fraud in the field: An experiment in public transportation. *Management Science*, (forthcoming).

Denton, K., & Krebs, D. (1990). From the scene to the crime: The effect of alcohol and social context on moral judgment. *Journal of Personality and Social Psychology, 59*(2), 242-248.

Dreber, A., & Johannesson, M. (2008). Gender differences in deception. *Economics Letters, 99*(1), 197-199.

Duke, A. A., & Bègue, L. (2015). The drunk utilitarian: Blood alcohol concentration predicts utilitarian responses in moral dilemmas. *Cognition, 134*, 121-127.

Erat, S., & Gneezy, U. (2012). White lies. *Management Science, 58*(4), 723-733.

Felson, R., Savolainen, J., Aaltonen, M., & Moustgaard, H. (2008). Is the association between alcohol use and delinquency causal or spurious? *Criminology, 46*(3), 785-808.

Fillmore, M. T., Carscadden, J. L., & Vogel-Sprott, M. (1998). Alcohol, cognitive impairment and expectancies. *Journal of Studies on Alcohol, 59*(2), 174-179.

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

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise–An experimental study on cheating. *Journal of the European Economic Association, 11*(3), 525-547.

Fligner, M. A., & Policello, G. E. (1981). Robust rank procedures for the Behrens-Fisher problem. *Journal of the American Statistical Association, 76*(373), 162-168.

Foerster, A., Pfister, R., Schmidts, C., Dignath, D., & Kunde, W. (2013). Honesty saves time (and justifications). *Frontiers in Psychology, 4*(473).

Fosgaard, T. R., Hansen, L. G., & Piovesan, M. (2013). Separating Will from Grace: An experiment on conformity and awareness in cheating. *Journal of Economic Behavior & Organization, 93*, 279-284.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25-42.

Friesen, L., & Gangadharan, L. (2012). Individual level evidence of dishonesty and the gender effect. *Economics Letters, 117*(3), 624-626.

Gächter, S., & Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature, 531*(7595), 496-499.

Galizzi, M. M., & Miraldo, M. (2017). Are you what you eat? Healthy behaviour and risk preferences. *The BE Journal of Economic Analysis & Policy, 17*(1).

Gino, F., & Ariely, D. (2012). The dark side of creativity: Original thinkers can be more dishonest. *Journal of Personality and Social Psychology, 102*(3), 445-459.

Gino, F., Ayal, S., & Ariely, D. (2013). Self-serving altruism? The lure of unethical actions that benefit others. *Journal of Economic Behavior & Organization, 93*, 285-292.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review, 95*(1), 384-394.

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *Quarterly Journal of Economics, 112*(2), 631-645.

Greene, J. D., & Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences, 106*(30), 12506-12511.

Gunia, B. C., Wang, L., Huang, L., Wang, J., & Murnighan, J. K. (2012). Contemplation and conversation: Subtle influences on moral decision making. *Academy of Management Journal, 55*(1), 13-33.

Gylfason, H. F., Arnardottir, A. A., & Kristinsson, K. (2013). More on gender differences in lying. *Economics Letters, 119*(1), 94-96.

Hanna, R., & Wang, S.-Y. (2013). Dishonesty and selection into public service. *NBER Working Paper Series, No. 19649*.

Hopthrow, T., Abrams, D., Frings, D., & Hulbert, L. G. (2007). Groupdrink: The effects of alcohol on intergroup competitiveness. *Psychology of Addictive Behaviors, 21*(2), 272-276.

Horrace, W. C., & Oaxaca, R. L. (2006). Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters, 90*(3), 321-327.

Jiang, T. (2013). Cheating in mind games: The subtlety of rules matters. *Journal of Economic Behavior & Organization, 93*, 328-336.

Jones, D. N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3). *Assessment, 21*(1), 28-41.

Kajackaite, A., & Gneezy, U. (2015). Lying costs and incentives. *Mimeo*.

Kocher, M. G., Schudy, S., & Spantig, L. (2017). I lie? We lie! Why? Experimental evidence on a dishonesty shift in groups. *Management Science*, (forthcoming).

Lane, S. D., Cherek, D. R., Pietras, C. J., & Tcheremissine, O. V. (2004). Alcohol effects on human risk taking. *Psychopharmacology, 172*(1), 68-77.

Levine, E. E., & Schweitzer, M. E. (2014). Are liars ethical? On the tension between benevolence and honesty. *Journal of Experimental Social Psychology, 53*, 107-117.

Levine, E. E., & Schweitzer, M. E. (2015). Prosocial lies: When deception breeds trust. *Organizational Behavior and Human Decision Processes, 126*, 88-106.

Lohse, T., & Qari, S. (2014). Gender differences in deception behaviour–The role of the counterpart. *Applied Economics Letters*, *21*(10), 702-705.

MacDonald, T. K., Zanna, M. P., & Fong, G. T. (1995). Decision making in altered states: Effects of alcohol on attitudes toward drinking and driving. *Journal of Personality and Social Psychology, 68*(6), 973-985.

Mann, S. A., Vrij, A., Leal, S., Vernham, Z. L., & Geven, L. (2015). The deterrence of deception through imposing cognitive load. *Open Access Journal of Forensic Psychology, 7*, 90-105.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research, 45*(6), 633-644.

Morewedge, C. K., Krishnamurti, T., & Ariely, D. (2014). Focused on fairness: Alcohol intoxication increases the costly rejection of inequitable rewards. *Journal of Experimental Social Psychology, 50*, 15-20.

Moskowitz, H., Burns, M. M., & Williams, A. F. (1985). Skills performance at low blood alcohol levels. *Journal of Studies on Alcohol, 46*(6), 482-485.

Muehlheusser, G., Roider, A., & Wallmeier, N. (2015). Gender differences in honesty: Groups versus individuals. *Economics Letters, 128*, 25-29.

Norton, E. C., Wang, H., & Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. *Stata Journal, 4*(2), 154-167.

Peterson, J. B., Rothfleisch, J., Zelazo, P. D., & Pihl, R. (1990). Acute alcohol intoxication and cognitive functioning. *Journal of Studies on Alcohol, 51*(2), 114-122.

Potters, J., & Stoop, J. (2016). Do cheaters in the lab also cheat in the field? *European Economic Review, 87*, 26-33.

Proestakis, A., Espín, A. M., Exadaktylos, F., Cortés Aguilar, A., Oyediran, O. A., & Palacio, L. A. (2013). The separate effects of self-estimated and actual alcohol intoxication on risk taking: A field experiment. *Journal of Neuroscience, Psychology, and Economics, 6*(2), 115-135.

Pruckner, G. J., & Sausgruber, R. (2013). Honesty on the streets: A field study on newspaper purchasing. *Journal of the European Economic Association, 11*(3), 661-679.

Rivers, D., & Vuong, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics, 39*(3), 347-366.

Ruffle, B., & Tobol, Y. (2017). Clever enough to tell the truth. *Experimental Economics, 20*(1), 130-155.

Sayette, M. A. (2017). The effects of alcohol on emotion in social drinkers. *Behaviour Research and Therapy, 88*, 76-89.

Schilbach, F. (2015). Alcohol and self-control: A field experiment in India. *Mimeo*.

Schreiber Compo, N., Evans, J. R., Carol, R. N., Kemp, D., Villalba, D., Ham, L. S., & Rose, S. (2011). Alcohol intoxication and memory for events: A snapshot of alcohol myopia in a real-world drinking scenario. *Memory, 19*(2), 202-210.

Shalvi, S. (2012). Dishonestly increasing the likelihood of winning. *Judgment and Decision Making, 7*(3), 292-303.

Shalvi, S., & De Dreu, C. K. (2014). Oxytocin promotes group-serving dishonesty. *Proceedings of the National Academy of Sciences, 111*(15), 5503-5507.

Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological Science, 23*(10), 1264-1270.

Sjöberg, L. (1969). Alcohol and gambling. *Psychopharmacology, 14*(4), 284-298.

Stepner, M. (2014). *Binned Scatterplots: Introducing binscatter and exploring its applications.* Paper presented at the 2014 Stata Conference.

Sutter, M. (2009). Deception through telling the truth?! Experimental evidence from individuals and teams. *The Economic Journal, 119*(534), 47-60.

van 't Veer, A., Stel, M., & van Beest, I. (2014). Limited capacity to lie: Cognitive load interferes with being dishonest. *Judgment and Decision Making, 9*(3), 199-206.

Weisel, O., & Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences, 112*(34), 10651-10656.

Wiltermuth, S. S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes, 115*(2), 157-168.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (2. ed.). Cambridge, Massachusetts: MIT Press.

Strategic Gaze: An Interactive Eye-Tracking Study

BY J. HAUSFELD, K. HESLER, AND S. GOLDLÜCKE

We present an interactive eye-tracking study that explores the strategic use of gaze. We analyze gaze behavior in an experiment with four simple games. The game can either be a competitive (hide & seek) game in which players want to be unpredictable, or a game of common interest in which players want to be predictable. Gaze is either transmitted in real time to another subject, or it is not transmitted and therefore non-strategic. We find that subjects are able to interpret non-strategic gaze, obtaining substantially higher payoffs than subjects who did not see gaze patterns. If gaze is transmitted in real time, eye-movements become more informative in the common interest games and players predominantly succeed to coordinate on efficient outcomes. In contrast, eye-movements become less informative in the competitive game.

JEL Codes: *C91, D01, D83, D87, C72*

Keywords: *Eye-tracking, focal points, signaling, hide & seek*

I. Introduction and Related Literature

In 1960, Thomas Schelling described a hypothetical experimental setting to test his theories of bargaining and focal points as follows: "The first stage in the experiment is to invent a machine, perhaps on the principle of the lie detector, that will record or measure a person's recognition or the focus of his attention […]" (Schelling, 1960, p. 109). Since then, eye-tracking, which allows to record where a person is looking, has developed into an affordable and precise method to analyze decision-making processes.[1] In particular, eye-tracking has proven useful to study how people play games and to infer the type of a player from recorded gaze data (Knoepfle, Wang & Camerer, 2009; Wang, Spezio & Camerer, 2010; Polonio, Di Guida & Coricelli, 2015; Stewart et al., 2016).[2] In contrast to this line of research, we extend the use of this technology to being an active component of the strategic interaction. Comparing strategic and non-strategic gaze allows us to explore how much communication—voluntary or involuntary—is possible with eye-movements.

Since eyes are a very natural means of communication, observing the eye-movements of a player should help to coordinate in common interest games. At the same time, it is obvious that eye-movements cannot convey too complex information.[3] In contrast, if players have an incentive to be unpredictable or even to deceive the other player, they should try not to let their gaze reveal their intentions. It could nevertheless be the case that "eyes don't lie" and that gaze always conveys information even if this is not in the players' interest. After all, professional poker players often wear sunglasses, eye-movements of amateur blackjack players can reflect the numerical value of their hands (Holmes, Ayzenberg & Lourenco, 2016) and it is very difficult to prevent recognition of previously seen faces from being shown in the eyes (Schwedes & Wentura, 2012). In other contexts, the sincerity of eye-movements is an illusion that might be exploited by liars. For example, many people, even professional interrogators, mistakenly believe that deception is associated with evasive eye-movements (e.g., Vrij, 2004).

---

[1] See, for example, Lahey & Oxley (2016) for a recent discussion on how this technology can be used to study economic theory.

[2] Also see the related literature using mouse-tracking like Costa-Gomes, Crawford & Broseta (2001) and Brocas et al. (2014).

[3] The limits of eye-to-eye communication are nicely expressed in the science fiction novel "The Dark Forest" by Liu (2008), in which two of the characters try to convey a message through facial expressions: "They stared at each other, but held that pose for less than half a minute before they burst out laughing at practically the same instant. 'My message was, `To-night I'd like to invite you to have supper on the Champs-Elysees,'' he said. She doubled over with laughter. 'Mine was, `You ... need to shave!''" (p. 179).

Recent research also explored how gaze can help in problem solving tasks. Several studies found that showing the recorded gaze patterns of successful problem solvers or experts in the task at hand to other participants can increase their task performance (see, e.g., Stein & Brennan, 2004; Litchfield et al., 2010; Litchfield & Ball, 2011). Brennan et al. (2008) and Neider et al. (2010) tested how bilateral live gaze transmission affects performance in different visual search tasks. They find that people are both better and faster in spotting the target in such tasks when they can see each other's gaze than when they cannot. It has further been shown that people are generally better at interpreting dynamic rather than static representations of gaze in various types of tasks (see, e.g., Nalanagula, Greenstein & Gramopadhye, 2006; Gallagher-Mitchell, Simms & Litchfield, 2017; Wermeskerken, Litchfield & Gog, 2017). Velichkovsky (1995) finds that an expert's verbal support in a puzzle solving task leads to even better performance when also his gaze is shown to the subject working on the task.[4]

In a closely related paper, Foulsham & Lock (2015) asked subjects to choose their most preferred out of four pictures (randomly generated, colorful fractals) and later showed the recorded gaze data to other subjects. While a large majority (~62%) of subjects were able to correctly guess which fractal was the favorite one from the recorded gaze data, this was no longer true when the eye-tracked subjects had been instructed to hide their true preference with their eye-movements (~28%).[5]

Our setting differs from the above mentioned studies. Participants were placed in a truly interactive environment and the strategic situation creates the incentive

[4] Müller et al. (2013) and Müller, Helmert & Pannasch (2014) showed that the benefits of gaze transfer in such tasks also have limitations, for instance, because the expert's gaze often also reflects his own search behavior instead of his intended advice such that the meaning of transmitted gaze can be ambiguous.

[5] Note, however, that the subjects trying to guess the preferred picture were not made aware that the choosing subject might have been trying to hide their true preference. Also, since the eye-tracked subjects were not incentivized to report their true preference, they could also have tried to mislead by simply not choosing their favorite picture.

what to communicate. We use eye-trackers connected to the software z-tree (Fischbacher, 2007) such that live gaze data can be displayed and integrated into the strategic interaction. In particular, in two of our treatments, one of the players sees a real-time representation of the other player's eye-movements on her own screen before having to choose. This technology opens up new possibilities to study strategic interactions. More specifically, we ask whether subjects strategically signal and deceive other subjects with their eyes in four simple 4×4 games when their focus of attention is shown to the other player on screen.

To assess the feasibility and limits of signaling one's intentions through gaze, we use three common interest games in which being predictable is in the interest of the eye-tracked player. Two of them are coordination games (one with and one without a focal option) in which subjects have to make the same choice in order to get a positive payoff. The third game is a discoordination game in which subjects earn money if they choose different options but not if they choose the same. Finally, we also explore if and how people utilize gaze transmission to conceal their true intentions in a hide & seek game in which being predictable would be costly for the eye-tracked player (the hider).

In order to investigate differences between strategic and non-strategic gaze, we further ran a treatment in which gaze was recorded but not transmitted to the other player. This non-strategic gaze was then shown to participants in another treatment who played against the decisions of the subjects in the treatment without gaze transmission.

We find strong evidence that people are able to correctly interpret non-strategic gaze patterns, both in the common interest games and in the hide & seek game. However, because non-strategic gaze is less focused in the hide & seek game, the effect is less pronounced there. When gaze is live transmitted to the other player, subjects can increase their success rates even further in the two coordination games but not in the discoordination game. It seems that the slightly more complex nature of this game makes it harder for the eye-tracked subject to commit to a clear gazing strategy. Finally, in the hide & seek game, we identify two general types of strategic gaze: Most people try to mask their intended choice with an uninformative gaze pattern while some try to actively mislead their counterparts into holding a wrong belief. Hiding intentions works almost perfectly in the sense that they are correctly detected at rates close to the random 25%. On average, actively trying to mislead does not pay, but we do find some evidence that it does work under specific circumstances. In general, hiders tend to gain an advantage when they do not choose the option they emphasized with their gaze, yet some of them nevertheless do the opposite.[6]

The findings from the hide & seek game are also applicable to sports in which players must try to anticipate their opponents next action. A prime example are penalty kicks in soccer. Tay et al. (2010) find that goal keepers are worse at predicting the direction of a shot when the kicker was instructed to try and deceive the goalie, e.g., by looking in the opposite direction before taking the shot.[7]

Our results further add to a recent literature that sheds light on the ability of people to predict what players will do from their facial expressions (Kovács-Bálint, Bereczkei & Hernádi, 2013; Van Leeuwen et al., 2017) or response times (Frydman & Krajbich, 2017). The questions how live gaze transmission in strategic interactions changes eye movements and how others will infer motives from these gaze patterns have not yet been addressed and are likely to gain importance in the digital age.

The remainder of the paper is structured as follows: We describe the experimental design and the games in the next section. In section III we briefly discuss

[6] From a game theoretic perspective, eye-movements should be seen as cheap talk and should thus reveal nothing in equilibrium. Our finding that too many of the misleaders nevertheless reveal their choice is related to Wang et al. (2010). They find that people also often over-communicate in cheap talk sender-receiver games with partially aligned interests such that the senders' true types can be inferred from their messages at a higher rate than random.

[7] Dicks, Button & Davids (2010) find a similar effect and additionally show that it is driven by the visual information in the earlier phases of the run-up before the kick. Also note however, that the performance of kickers was found to decrease when they spend less time focusing their actual target (Nagano, Kato & Fukuda, 2006).

our two central hypotheses concerning gaze and choice behavior. Our results are presented in section IV and we conclude in section V.

II. Experimental Design

A. Games

All participants played four different two-player games: A coordination game, a focal coordination game, a focal discoordination game, and a hide & seek game. Each of the games was played for five consecutive rounds with random re-matching of participants and no feedback between rounds. In all 20 rounds, both players saw four large, gray rectangles arranged in a square on their computer screens and it was their task to pick one of them.[8] Their earnings were determined by whether or not the two players' actions matched: In the coordination and the focal coordination game, both participants only earned points when they chose the same box. In the focal discoordination game, both participants only earned points when their actions did *not* match. In the hide & seek game, the hider would earn points if the actions did not match whereas the seeker would earn points if the actions did match.[9]

The maximum number of points a player could earn by choosing a certain box was always displayed in the center of the box. In the coordination game, the maximum number of points was always equal to 10 in all four boxes, i.e., both players earned 10 points if they managed to coordinate on one box and received 0 otherwise. In the case of the focal coordination game, only the top left box gave 10 points whereas the other three boxes all gave 11 points if the players

---

[8] Subjects made their choices by pressing one of four buttons on a normal computer keyboard which were marked with stickers. The buttons used were R, U, C, and N which approximately mirror the positioning of the boxes on screen.

[9] We counterbalanced the order of games within each session. It was either 1. hide & seek, 2. coordination, 3. focal discoordination, and 4. focal coordination or 1. coordination, 2. hide & seek, 3. focal coordination, and 4. focal discoordination.

managed to coordinate on them.[10] Similarly, in the focal discoordination games, if the players managed to choose different boxes, any box apart from the top left one (which yielded 10 points) earned the player who chose it 11 points. If they chose the same box in the focal discoordination games, both would earn 0 points. Finally, in the hide & seek game, always one of the players earned 10 points—the seeker if their actions matched and the hider if they did not. Figure 1 shows the four games in normal form.[11]

Our participants entered their decisions sequentially, i.e., the eye-tracked player always chose first while the other had to wait. The boxes were already displayed to the second mover as she was waiting. Once the first mover had entered his decision, a green frame appeared around the boxes on the second mover's screen indicating to her that she could now enter her decision. Before each round, subjects saw a brief reminder of the payoff consequences of matching and non-matching choices for both players for eight seconds, followed by a two second display announcing the start of the next round, and a finally a fixation cross in the center of the screen for one second.

	TL	TR	BL	BR
TL	10,10	0,0	0,0	0,0
TR	0,0	10,10	0,0	0,0
BL	0,0	0,0	10,10	0,0
BR	0,0	0,0	0,0	10,10

(a) Coordination Game

	TL	TR	BL	BR
TL	10,10	0,0	0,0	0,0
TR	0,0	11,11	0,0	0,0
BL	0,0	0,0	11,11	0,0
BR	0,0	0,0	0,0	11,11

(b) Focal Coordination Game

	TL	TR	BL	BR
TL	0,0	10,11	10,11	10,11
TR	11,10	0,0	11,11	11,11
BL	11,10	11,11	0,0	11,11
BR	11,10	11,11	11,11	0,0

(c) Focal Discoordination Game

	TL	TR	BL	BR
TL	0,10	10,0	10,0	10,0
TR	10,0	0,10	10,0	10,0
BL	10,0	10,0	0,10	10,0
BR	10,0	10,0	10,0	0,10

(d) Hide & Seek Game

FIGURE 1. THE FOUR GAMES IN NORMAL FORM.

Notes: The numbers in the cells correspond to the players' respective payoffs: The first number always shows the row-player's (hider's) payoff. The second number shows the column-player's (seeker's) payoff. The labels TL, TR, BL, and BR represent the top left, top right, bottom left, and bottom right box on the participants' computer screens, respectively.

[10] Similar games were also used by Bardsley et al. (2010) to study how focal points are used for coordinating on an equilibrium in one-shot games.

[11] See Appendix A for a list of all equilibria but note that all possible individual choices are consistent with Nash equilibrium in all games.

B. Treatments

We used four treatments to explore the role of gaze: *NoGaze, RecordedGaze, LiveGaze-FreeChoice,* and *LiveGaze-ForcedChoice*. In the *NoGaze* treatment, the gaze of the first mover was recorded but it was not shown to the second mover.[12] With this treatment, we can investigate the gaze patterns of non-strategic gaze and the behavioral results further serve as a benchmark without communication. In the *RecordedGaze* treatment, we displayed this earlier recorded non-strategic gaze data to subjects who then also played against the corresponding old decisions. This allows us to infer whether and when subjects understand non-strategic gaze. In the two remaining treatments, the first mover's gaze is transmitted in real time to the second mover. In the *LiveGaze-FreeChoice* treatment, the first mover could freely choose any of the four boxes in all games and would earn points according to the payoffs shown in Figure 1. This was not the case in the *LiveGaze-ForcedChoice* treatment. Here, the first mover could still opt for any box of his choice but would only earn points if the chosen box coincided with the one that was randomly selected by the computer beforehand. This box was marked by an arrow which was only visible to him (and not to the second mover) for two seconds before the fixation cross appeared. In these two treatments, gaze can be used strategically as a device for communication. In the *ForcedChoice* treatment, subjects can focus on choosing a good gazing-strategy as the only meaningful box-choice is already indicated by the computer.[13]

The gaze data was visualized by changing the color of the box that was currently looked at by the first mover from gray to red (see Figure 2). This was the case both on the screen of player 1 and on that of player 2, i.e., player 1 was

[12] We used eye trackers with a sampling rate of 60 Hz. Subjects were seated approximately 58 centimeters in front of their computer screens, all of which were of the same model and had a resolution of 1280×1024 pixels. To improve data quality, we also used chin rests for all eye-tracked subjects.

[13] Only seven subjects did not always choose the box selected by the computer, four of which did so once and the remaining three did so on several occasions (in two, five, and eight out of the 20 games, respectively).

fully aware of how his gaze data was displayed to player 2. Although no gaze data was transmitted between players in the *NoGaze* treatment, we nevertheless kept the treatment as symmetric as possible: For the first movers, the boxes still changed color according to where they looked just like in the *LiveGaze* treatments. Likewise, second movers also had to wait for the first mover to enter his decision before she could enter hers—she merely did not see any gaze data.

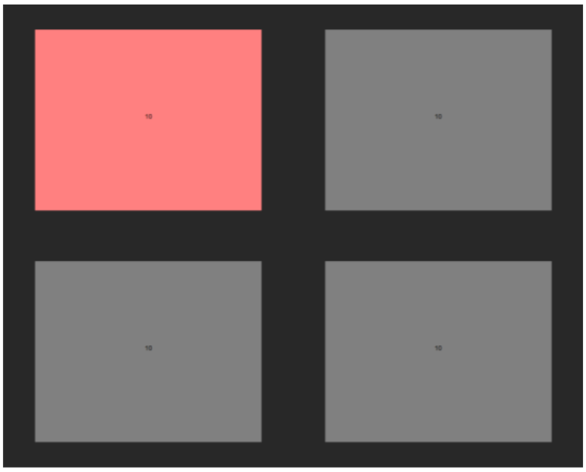


FIGURE 2. SCREENSHOT OF A SUBJECT CURRENTLY LOOKING AT THE TOP LEFT BOX.

All participants in the *RecordedGaze* treatment played against the exact same sequence of recorded first movers as was actually faced by one of the second movers in the *NoGaze* treatment (recall that the groups were randomly rematched after each round).

To account for our method of gaze visualization in the analysis, we deviate from common fixation definitions which typically require that the position of gaze remains constant for some minimum amount of time (see, e.g., Salvucci & Goldberg, 2000). Instead, we consider everything a fixation that was actually displayed to the participants. A fixation thus starts the moment the respective box turns red and lasts until the box turns gray again.

We ran three sessions in each treatment with eye-tracking (*NoGaze*, *LiveGaze-FreeChoice, and LiveGaze-ForcedChoice*), each with 24 subjects of which the twelve first movers were eye-tracked. Data acquisition for the *RecordedGaze* treatment was finished within one session with 30 second movers.[14] The experiment was programmed in z-Tree (Fischbacher, 2007) and our subjects were recruited from the general student population of the University of Konstanz using the software ORSEE (Greiner, 2004). Each session lasted around one hour and subjects were paid €0.1 for each point they earned in the experiment such that they received an average of around €17. Our participants had an average age of about 21 years and about 40% of them were male. A translation of the instructions given to the participants as well as further details on the experimental procedures can be found in Appendix B. Except for the different treatments, all features of the experiments were made common knowledge to the participants.

III. Hypotheses

Our two central hypotheses are based on the conjecture that gaze can be used as a communication device to announce a strategy choice (e.g., by focusing on one of the boxes for a long time).

HYPOTHESIS 1. INTERPRETING GAZE

Subjects are able to correctly infer choice in strategic games from the observed gaze patterns of the choosing subject.

There exists a larger literature that shows that decisions often follow eye-movements in the sense that the chosen option is both the one that was looked at last and for a longer time (see, e.g., the gaze cascade effect as described by Shimojo et al., 2003, or the attentional drift diffusion model by Krajbich, Armel & Rangel, 2010). As extensively reviewed in the introduction, there are numerous studies showing that people's problem solving performance can benefit from seeing the eye movements of others who previously worked on the same

[14] Note that there were 36 second movers in the *NoGaze* treatment which all faced a different random sequence of first movers. Out of these 36 sequences, we randomly selected 30 to be faced by the 30 second movers in the *RecordedGaze* treatment.

task (e.g., Stein & Brennan, 2004; Nalanagula et al., 2006; Litchfield et al., 2010; Litchfield & Ball, 2011).

We can test whether subjects can exploit the information of others' gaze also in strategic games by comparing the success rates of the second movers in the *NoGaze* treatment with those in the other treatments. We can further test the limits of this ability on two dimensions. First, we vary the nature of the strategic interaction between the games (common interest vs competitive) and second, we vary whether subjects have an incentive to affect the informativeness of their gaze (recorded gaze vs live gaze transmission). Our second hypothesis is concerned with how the first movers react to this incentive.

HYPOTHESIS 2. STRATEGIC GAZING

Subjects are aware that their gaze conveys information and are able to strategically adapt it to their advantage.

If people are able to infer meaning from another person's eye movements, then there are obvious implications for strategic interactions. If being predictable is in a player's self-interest, then she should of course try to communicate as clearly as possible.[15] If, however, being predictable is not in the player's self-interest, she should obviously try to remove all meaning from any signal she sends out. Wang et al. (2010) show that people often fail at keeping their true intentions hidden from the messages they send in sender-receiver games. There further exists a plethora of studies in the field of sport psychology that investigate how an athlete's actions can be predicted from cues observable shortly before the action (often including the direction of gaze). Many of which also showed that such cues can be mimicked in order to deceive the opponent.[16]

[15] Note that, in the common interest games, communicating one's intended choice is both "self-committing" and "self-signaling". These properties ensure that communicating intentions is in a player's interest if and only if she indeed makes the indicated choice. They make an informative equilibrium more likely than the babbling equilibrium in which all messages are ignored (see, e.g., Farrell & Rabin, 1996).

[16] Examples include studies from soccer (e.g., Tay et al., 2010), tennis (e.g., Rowe et al., 2009), rugby (Jackson, Warren & Abernethy, 2006), basketball (Sebanz & Shiffrar, 2009), or handball (Cañal-Bruland & Schmidt, 2009).

We can test if and how successfully people make use of the opportunity to strategically adapt their eye movements by comparing gaze patterns and success rates across our treatments. The three different common interest games further allow us to explore the limits of the ability to send clear signals with one's eyes. For instance, live gaze could help avoiding the focal but inefficient top left option in the focal games. The *LiveGaze-ForcedChoice* treatment additionally allows us to investigate whether people become better at strategic gazing when they do not have to concentrate on making a choice as the computer chooses randomly for them. Finally, the hide & seek game allows us to explore how subjects strategically adapt their gaze when it is in their interest to be unpredictable. We investigate two straight forward candidate strategies: Signal-jamming, i.e., hiding all information from one's gaze or active misleading, i.e., trying to trick the opponent into holding a wrong belief.

IV. Results

We begin our analysis with an overview of the average choice data. The choice data from the *NoGaze* treatment can be used to explore whether there is some bias toward any of the four boxes. For instance, a western reading style would order the boxes from left to right and top to bottom such that the top left box could be considered as "first" and could therefore be perceived as salient (see, e.g., Abed, 1991; Chua, Boland & Nisbett, 2005; and Ishii et al., 2011, for studies on this reading bias). Table 1 shows the choice distribution of both first and second movers in all treatments with free choice.[17] Despite the symmetric layout of the four boxes in a square, the choice data from the *NoGaze* treatment reveal a tendency of the first movers to choose the top right box particularly often (47% trials, a χ^2 test of the hypothesis that all boxes are chosen with equal probabilities is clearly rejected with $p < .0001$). This bias is not reflected in the

[17] Since the vast majority of first mover choices in the *LiveGaze-ForcedChoice* treatment complied with the random choice made by the computer (only 2.6% deviations), the choice distribution is not telling in that case.

second movers' decisions though: They rather display a bias favoring both top boxes but not any of the two in particular (χ^2 test, $p < .01$). Since second movers had to wait for the first movers' input also in the *NoGaze* treatment, a potential reason for this difference between first and second movers is the mere timing.[18]

Compared to the findings by Bardsley et al. (2010), the focal option in the focal coordination game was used less frequently as a coordination device, especially so by first movers who only chose it in 35% of the trials compared to 59% of the second movers. When second movers could see the previously recorded gaze, their choice distribution seems to harmonize more with that of the first movers. This effect seems to become even stronger with live gaze transmission under which the focal but inefficient outcome is now avoided almost completely by both players.

In the focal discoordination game, the focal option is chosen least often by the first movers, especially so in the *LiveGaze-FreeChoice* treatment (χ^2 test marginally significant in the *NoGaze* treatment with $p = .058$ and highly significant in *LiveGaze-FreeChoice* with $p < .001$). Except in the *RecordedGaze* treatment, also second movers' choice distributions in the focal discoordination game are significantly different from random ($p > .15$ in *RecordedGaze* and $p < .02$, else). Our results from the *NoGaze* treatment slightly differ from the results by Rubinstein, Tversky & Heller (1997) who find that people's choice distributions in most discoordination games did not significantly differ from random play.[19]

[18] Penczynski (2016) finds evidence for such effects even when there is no actual difference in timing but when there merely are different framings that put either the hider or the seeker in the role of the first mover (albeit they always decide simultaneously).

[19] They too had a focal option in each game. However, here salience was induced by using different labels rather than different payoffs (e.g., by using "A, B, A, A" as labels).

TABLE 1—AVERAGE BOX CHOICES IN TREATMENTS WITH FREE CHOICE.

	NoGaze Treatment			
	Coordination	Focal Coordination	Focal Discoordination	Hide & Seek
Top Left	24% (31%)	35% (59%)	18% (27%)	22% (24%)
Top Right	47% (32%)	38% (18%)	23% (17%)	17% (29%)
Bottom Left	9% (21%)	9% (11%)	32% (28%)	32% (21%)
Bottom Right	19% (17%)	17% (12%)	27% (28%)	29% (26%)

	RecordedGaze Treatment			
	Coordination	Focal Coordination	Focal Discoordination	Hide & Seek
Top Left	27% (27%)	35% (25%)	18% (17%)	23% (27%)
Top Right	44% (41%)	38% (42%)	23% (27%)	18% (19%)
Bottom Left	11% (13%)	9% (11%)	34% (27%)	31% (30%)
Bottom Right	19% (20%)	18% (21%)	25% (29%)	29% (25%)

	LiveGaze-FreeChoice Treatment			
	Coordination	Focal Coordination	Focal Discoordination	Hide & Seek
Top Left	34% (36%)	1% (1%)	12% (10%)	26% (29%)
Top Right	37% (36%)	46% (46%)	34% (23%)	16% (27%)
Bottom Left	13% (12%)	17% (18%)	29% (35%)	34% (24%)
Bottom Right	17% (16%)	36% (36%)	24% (32%)	23% (20%)

Notes: The table shows the choice distributions of the eye-tracked participants and those of the second movers in parentheses. 36 first and second movers in the *NoGaze* and the *LiveGaze-FreeChoice* treatments, 30 second movers in the *RecordedGaze* treatment. Data of first movers in the *RecordedGaze* treatment stem from the 30 randomly selected first movers from the *NoGaze* treatment. Each participant played each game five times. Totals unequal 100% are due to rounding.

Finally, in the hide & seek game, first movers in the *NoGaze* treatment again choose significantly different from random (χ^2 test, *p*-value < .05) whereas the second movers cannot be distinguished from random play (*p* > .5). Note however, that the first movers' top right bias observed in the coordination game now disappeared and instead the bottom boxes are chosen more frequently. The choice distribution of second movers in the *RecordedGaze* treatment is now much closer to that of the first movers and also significantly different from random (*p* < .05). Interestingly, in the *LiveGaze* treatment, the first movers again exhibit a slight bias favoring the bottom boxes (*p* < .01) but, unlike in *RecordedGaze*, the second movers now do not play significantly different from random like in *NoGaze* (*p* > .3).[20]

[20] Also see Rapoport & Budescu (1992) who find that subjects behave more randomly in matching pennies than when they are asked to behave "randomly". Nevertheless, the literature on sequential matching pennies and hide & seek games has shown that behavior is often consistent with equilibrium in the aggregate, but that choices are serially correlated on the individual level (see, e.g., Walker & Wooders, 2001; Levitt, List & Reiley, 2010; Emara et al., 2017).

In the following four sub-sections, we present a detailed analysis of the effects of gaze transfer. We first analyze whether and how subjects interpret non-strategic gaze. This is followed by three sub-sections on strategic gaze. The first is concerned with the effectiveness of strategic gaze in terms of matching rates in the different games which is followed by an analysis of the characteristics of strategic gaze. Finally, we analyze the different gaze strategies employed by our participants in the hide & seek game.

A. Non-Strategic Gaze

Figure 3 depicts the average success rates of the second movers in the *NoGaze* and the *RecordedGaze* treatments.[21] Participants in *RecordedGaze* were clearly able to correctly interpret the non-strategic gaze patterns of the first movers previously recorded in the *NoGaze* treatment. Second movers in the *RecordedGaze* treatment achieved significantly higher success rates in all four games (Wilcoxon rank-sum tests comparing individual average success rates in each game by treatment, all Bonferroni corrected p-values < 0.03). In the three common interest games, the second movers do remarkably well in interpreting the recorded gaze patterns. Though the success rate in the hide & seek game is also significantly higher when non-strategic gaze can be observed, the gained advantage seems to be less pronounced.[22]

---

[21] A trial is counted as a success for the second mover when she receives a positive payoff, i.e., when the two players choose the same box in the coordination, the focal coordination, and the hide & seek game and when they choose different boxes in the focal discoordination game.

[22] Also note that, in line with the findings of Rubinstein et al. (1997), we find a small seeker advantage also in the hide & seek games of the *NoGaze* treatment as the matching rate is slightly above the 25% expected in equilibrium. This was later ascribed to a “fatal attraction” to the “least salient option” by Crawford & Iriberri (2007) which is also in line with our results in the sense that the first movers in the hide & seek game most often choose the bottom left box which is the one they choose least often in the coordination game.

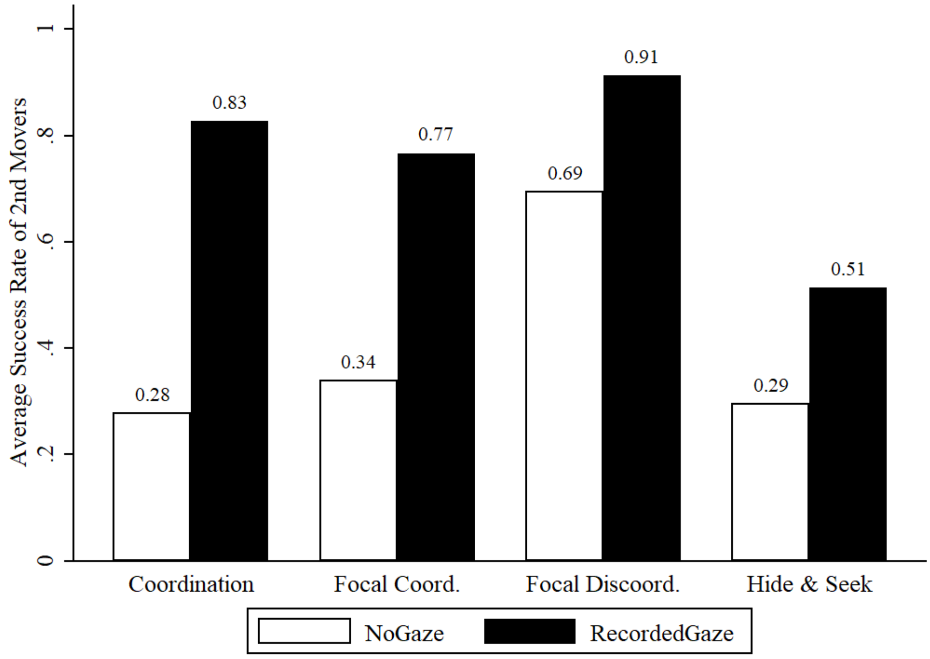


FIGURE 3. AVERAGE SUCCESS RATES OF SECOND MOVERS IN TREATMENTS NO GAZE AND RECORDED GAZE.

Notes: An outcome is counted as a success for the second mover if his payoff is nonzero, i.e., when he matches the eye-tracked subject's box choice in the (focal) coordination and hide & seek games and when he chooses a different box than the eye-tracked subject in the focal discoordination game.

The second movers' smaller gain from seeing the first movers' eye movements in the hide & seek game could be explained by differences in the non-strategic gaze patterns between the games. For instance, gaze from the hide & seek game could be more difficult to interpret if the first mover looks back and forth between boxes a lot or if the fixations on individual boxes are particularly short. However, neither the average number of transitions nor the average fixation duration differed significantly between the games (Wilcoxon signed ranks tests, all Bonferroni corrected p-values > 0.8).

How else could the difference be explained? In order to gain an advantage from seeing the first mover's gaze, the second mover has to come up with a theory of how the first mover's gaze translates into his choice. Motivated by the literature reviewed above, we compared the predictive success of three simple heuristics that second movers might resort to in order to interpret gaze.

The *Most Looked* heuristic predicts that the first mover chooses the box that he looks at for the longest time. The *Last Looked* heuristic states that he chooses the box that he looks at last. Finally, we consider a *Frugal Tree* heuristic (see Raab & Gigerenzer, 2015, for a recent discussion of frugal trees). A frugal tree features multiple stages at which a choice can be made. In each stage, a choice criterion is checked. If it is fulfilled, the corresponding choice is made and the process ends. If not, the process moves on to the next stage and checks the corresponding criterion. The frugal tree that we consider combines the two heuristics from above. It first checks the *Most Looked* heuristic but moves on to the *Last Looked* heuristic if the first criterion is not clear enough. This is the case when it becomes difficult to discern which box has actually been looked at the most, i.e., when the inspection time on the most and second most looked at boxes are not sufficiently different from one another. Formally, we require that the most looked at box was inspected at least twice as long as each of the other three boxes.[23]

Table 2 shows the average success rates the second movers in the *RecordedGaze* treatment would have achieved if they all had followed the respective heuristics.[24] Clearly, the heuristics perform much worse in the hide & seek game than in all others. This can explain why the second movers in the *RecordedGaze* treatment could profit more from seeing the gaze patterns in the common interest games than in the hide & seek game because gaze is less informative about the actual choice. These differences also turn out to be significant in pairwise

[23] Since we cannot know the exact criterion at which subjects switch from the *Most Looked* heuristic to the *Last Looked* heuristic, this parameter choice is somewhat arbitrary. In our sample, the highest predictive power across all games and treatments (80.5%) would be reached if we required that the most inspected box was inspected at least γ = 3.64 times as long as each of the other four. We nevertheless stick to this rule of thumb because of its simplicity and because the main increase in predictive power of the heuristic (from 76.1% to 80%) happens when γ is increased from 1 to 2. All further increases only improve the average success rate by very little.

[24] Table 11 in Appendix C shows the share of second movers that in fact decided in line with these heuristics. Most second movers indeed followed one of the heuristics.

Wilcoxon rank-sum tests comparing the success rates between games for each of the three heuristics.[25]

Apart from one exception, the success rates of the three heuristics do not significantly differ between the three common interest games (all corrected *p*-values > .1). Only the success rate of the *Last Looked* heuristic is significantly higher in the coordination game than in the focal discoordination game (corrected $p < .1$) in the *RecordedGaze* treatment. Among the four games, the coordination game is also the one in which the participants benefited the most from seeing the recorded gaze patterns. In summary, we find strong support for our first hypothesis: People are indeed able to correctly interpret gaze patterns, even when they are purely instrumental and not willingly controlled in order to influence their informativeness. Choice in the hide & seek game is harder to predict from the non-strategic gaze patterns than in the other games, but subjects could nevertheless significantly increase their success rate even there.

TABLE 2—SHARES OF CORRECT PREDICTIONS BY THE THREE HEURISTICS IN RECORDED GAZE TREATMENT.

	Coordination	Focal Coordination	Focal Discoord.	Hide & Seek
Most Looked	0.84	0.88	0.83	0.61
Last Looked	0.94	0.91	0.85	0.72
Frugal Tree	0.95	0.93	0.87	0.68

Notes: The table shows each heuristic's share of correct predictions in each game. This also corresponds to the second movers' success rate which they would have achieved if they had all followed the predictions of the respective heuristic.

B. Strategic Gaze: Success Rates

Figure 4 shows the average success rates of the second movers in each of the games in the *RecordedGaze* and the two *LiveGaze* treatments. In the two coordination games, we can see that subjects could increase their success rate even further when the first movers' gaze was live transmitted to their current partner.

[25] For the *Most Looked* heuristic and the *Frugal Tree*, all Bonferroni corrected *p*-values < 0.002. Also for the *Last Looked* heuristic, predictive success in the hide & seek game is highly significantly lower than in the (focal) coordination game (both corrected *p*-values < 0.002) but to a lesser extent in case of the focal discoordination game (corrected $p = 0.072$).

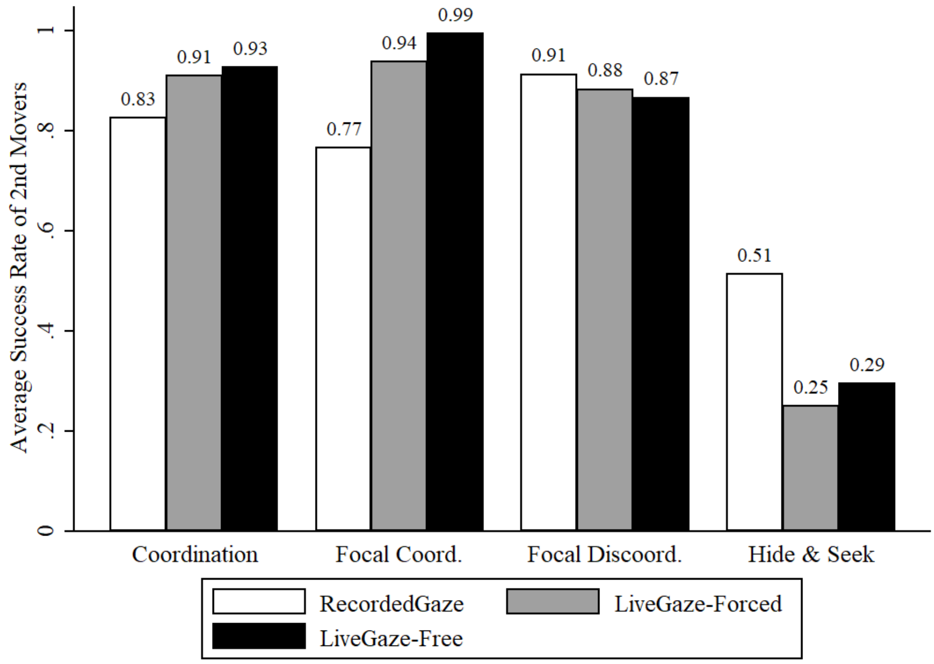


FIGURE 4. AVERAGE SUCCESS RATES OF SECOND MOVERS IN RECORDED AND LIVE GAZE TREATMENTS.

Notes: An outcome is counted as a success for the second mover if his payoff is nonzero, i.e., when he matches the eye-tracked subject's box choice in the (focal) coordination and hide & seek games and when he chooses a different box than the eye-tracked subject in the focal discoordination game.

In the coordination game, the success rate significantly increased from 83% to 91% and 93% in the *LiveGaze-ForcedChoice* and the *LiveGaze-FreeChoice* treatments (pairwise Wilcoxon rank-sum tests comparing individual average success rates by treatment, both corrected *p*-values < 0.04). In the focal coordination game, the effect is even more extreme: The success rate of 77% in the *RecordedGaze* treatment is increased to 94% in the *LiveGaze-ForcedChoice* and even to 99% in the *LiveGaze-FreeChoice* treatment (both corrected *p*-values < 0.04).[26] However, in the focal discoordination game, the success rates are actually slightly lower in the *LiveGaze* treatments than in the *RecordedGaze* treatment but these differences are far from significant (both *un*corrected *p*-values > .4). Despite the slight decrease in the focal discoordination game, the live

[26] This is in line with the bargaining experiment of Isoni et al. (2013), who show that communication is very effective in choosing efficient payoff divisions and salient features only matter for the decision between similarly efficient and equal outcomes.

gaze transmission nevertheless led to significantly higher success rates compared to the *NoGaze* treatment (both corrected *p*-values < 0.001). Thus, by using their gaze as a signaling device, first movers were clearly able to make themselves more predictable than without gaze transmission in all three common interest games.

The picture changes for the case of the hide & seek game. The success rates in the two *LiveGaze* treatments (25% and 29%) are both significantly smaller than in the *RecordedGaze* treatment (51%, both corrected *p*-values < 0.01), but do not differ significantly from the 29% in the *NoGaze* treatment (both *un*corrected *p*-values > 0.3). Hence, first movers were also able to *hide* their intentions when it was in their interest to become *less* predictable. However, they were not able to significantly increase their success rate compared to the *NoGaze* treatment.

C. Strategic Gaze: Characteristics

To get a first impression of how the first movers made use of the opportunity to gaze strategically, we again compared the average number of fixations and the average fixation durations across games (see Table 3). Unlike in the *RecordedGaze* treatment, the average number of fixations is significantly higher in the hide & seek game compared to the three common interest games in both *LiveGaze* treatments (pairwise Wilcoxon signed ranks tests, all corrected *p*-values < .02). Similarly, the average fixation duration is significantly shorter in the hide & seek game than in the other three games for both *LiveGaze* treatments (all corrected *p*-values < .01). In the hide & seek game, subjects thus moved their gaze more quickly from one box to the next and also did so more often than in the common interest games, making their gaze more difficult to interpret.

Apart from one exception, the gaze characteristics did not differ significantly between the three common interest games in neither treatment (all corrected *p*-values > .15). The exception lies in the *LiveGaze-ForcedChoice* treatment: Here, both the average number of fixations is greater and the average fixation duration is shorter in the focal discoordination game than in the focal coordination game (both corrected p-values < .08). This could also explain why the success rate in the focal discoordination game in the *LiveGaze* treatments did not improve compared to the *RecordedGaze* treatment.

TABLE 3—AVERAGE FIXATION NUMBERS AND FIXATION DURATIONS ACROSS GAMES AND TREATMENTS.

	Average Number of Fixations			
	Coordination	Focal Coordination	Focal Discoord.	Hide & Seek
RecordedGaze	4.6	5.3	5.3	5.3
LiveGaze-Forced	2.8	2.6	5.3	9.3
LiveGaze-Free	4.7	3.4	4.8	13.8
	Average Fixation Duration in Seconds			
	Coordination	Focal Coordination	Focal Discoord.	Hide & Seek
RecordedGaze	1.05	.83	1.00	.66
LiveGaze-Forced	2.57	2.67	1.84	.73
LiveGaze-Free	1.85	2.43	1.95	.67

Notes: The top half of the table shows the average number of fixations between boxes per game for each treatment. The bottom half shows the according average fixation durations in seconds. We consider everything a fixation that was displayed to the participants as a red box, i.e., a fixation starts when the respective box turns red and ends when it turns gray again.

These effects also become apparent in the predictive success of the three heuristics we introduced earlier (the heuristics' respective shares of correct predictions in all four games are shown in Table 4 below). In both *LiveGaze* treatments, the success rates of all three heuristics in the hide & seek game drop to values of 30% and below. This again clearly shows that first movers actively try to keep their gaze as uninformative about their intentions as possible. In contrast, in the coordination as well as the focal coordination game, the heuristics' hit rates lie between 94% and 100% indicating that in these games the first movers succeed in being particularly predictable. Finally, first movers seem to have more difficulties in sending clear signals in the focal discoordination game as the heuristics' success rates only lie between 74% and 81% there.[27]

The heuristics' hit rates in the focal coordination game in the *LiveGaze-FreeChoice* treatment are exceptionally high (99% for the *Last Looked* heuristic

[27] Success rates of all heuristics are significantly smaller in the hide & seek game than in all other games (Wilcoxon rank-sum tests, all corrected p-values < .002). Further, success rates in the focal discoordination game are significantly smaller than in the (focal) coordination games (all corrected p-values < .002), and they do not significantly differ between the focal and the non-focal coordination game (all corrected p-values > .1).

and even 100% for the other two). This indicates that the presence of the salient but inefficient outcome in the focal coordination game makes subjects exhibit the most extreme gaze patterns. This also becomes apparent in the sense that they clearly avoid looking at the focal option: Compared to the coordination game without a focal payoff in the top left box, the share of first fixations on the top left box drops from more than 50% to less than 15% in the focal coordination game. Similarly, the share of trials in which the top left box is not even looked at once increases from 36% to 71%.

TABLE 4—SHARES OF CORRECT PREDICTIONS BY THE THREE HEURISTICS IN LIVE GAZE TREATMENTS.

	LiveGaze-ForcedChoice			
	Coordination	Focal Coordination	Focal Discoord.	Hide & Seek
Most Looked	0.97	0.96	0.74	0.22
Last Looked	0.96	0.99	0.78	0.30
Frugal Tree	0.97	0.99	0.78	0.28

	LiveGaze-FreeChoice			
	Coordination	Focal Coordination	Focal Discoord.	Hide & Seek
Most Looked	0.96	1.00	0.79	0.26
Last Looked	0.94	0.99	0.79	0.30
Frugal Tree	0.97	1.00	0.81	0.30

Notes: The table shows each heuristic's share of correct predictions in each game for the *LiveGaze-ForcedChoice* (top) and the *LiveGaze-FreeChoice* treatment (bottom). This also corresponds to the second movers' success rate which they would have achieved if they had all followed the predictions of the respective heuristic.

In the focal discoordination game, the heuristics' hit rates are significantly lower than in the other common interest games. This indicates that the ability to send clear signals also depends on the nature of the strategic situation at hand. A discoordination game adds complexity to the situation as the signal now becomes ambiguous: It could both be an announcement of the first mover's own strategy but also a choice recommendation for the second mover. This aspect could make it harder for the first movers to commit to a particular gazing strategy such that their gaze patterns become more difficult to interpret. Also note that the share of second movers deciding in line with a belief based on the respective heuristics never drops below 90% also in the focal discoordination game (see Table 11 in Appendix C). This indicates that the failure to increase the success rate in the discoordination game with strategic compared to non-strategic gaze is in fact due to first movers not choosing the emphasized box.

In order to explore the effects of the strategic environment on the signal quality more closely, we also analyzed it in terms of the distribution of viewing time across the four boxes. In order to send a clear signal, one box should be emphasized as much as possible. This is achieved when the share of viewing time on all boxes *but* the most inspected box is minimized. Figure 5 shows the cumulative distribution of the share of viewing time which subjects allocated to all but the most inspected box (separately for all four games and the two *LiveGaze* treatments). The clearest possible signal—only looking at one single box—corresponds to a share of zero in the figure. This means that the intercept with the *y*-axis shows the share of games in which the respective first mover only looked at a single box.

This share is largest in the focal coordination game in both *LiveGaze* treatments (86% and 67%, respectively). In the *LiveGaze-FreeChoice* treatment, the cumulative distribution of gaze time in the focal coordination game is always to the left of that in the coordination game, i.e., gaze in the focal coordination game is particularly clear also with respect to the allocation of viewing time.[28]

This is different in the case of the focal discoordination game in which complexity is increased due to the ambiguity of the signal. In both treatments, the share of first movers that look at only one box is lower in the focal discoordination game than in the focal coordination game. The cumulative distributions of gaze time in the focal discoordination game are also always to the right of their counterparts from the focal coordination game and these differences are highly significant (Kolmogorov-Smirnov tests, both corrected p-values $< .001$). Signal strength in terms of viewing time thus clearly suffers in the focal discoordination game.

[28] According to a two-sided Kolmogorov-Smirnov test, this difference is also highly significant (corrected p-value < 0.001).

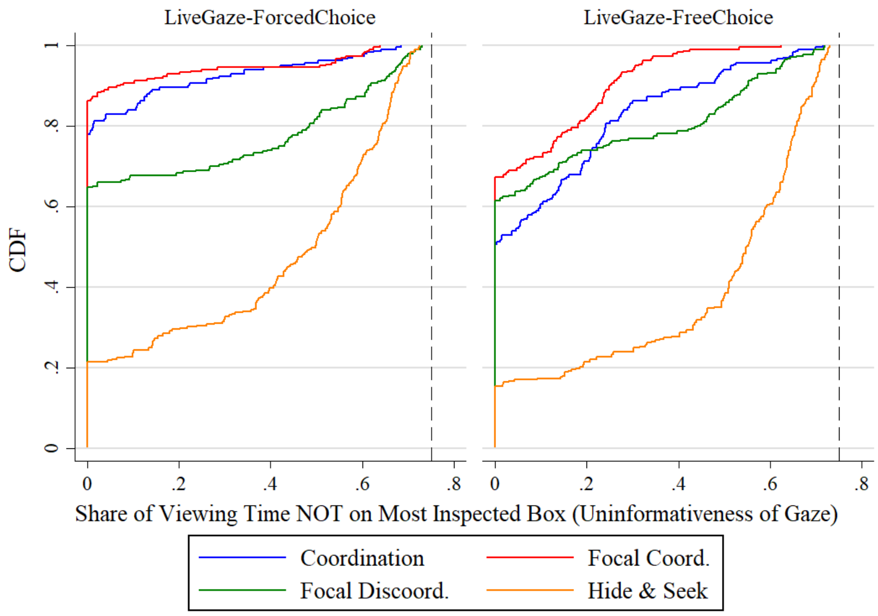


FIGURE 5. CUMULATIVE DISTRIBUTION OF SHARE OF VIEWING TIME ON ALL BOXES OTHER THAN THE MOST INSPECTED BOX ACROSS GAMES AND TREATMENTS.

Notes: The graphs show the distribution of the share of viewing time that the subjects allocated to all boxes except the one they inspected the most. The intercept with the *y*-axis thus corresponds to the percentage of people that only looked at one single box in the respective game and treatment. The dashed lines mark the natural maximum of 75% viewing time on the non-most inspected boxes which would be reached if all four boxes were inspected for exactly 25% of the time.

D. Strategic Gaze in Hide & Seek: Signal-Jammers and Misleaders

We now continue with an extensive analysis of the gaze patterns in the hide & seek games in which first movers have an incentive to be as unpredictable as possible. Note that in Figure 5, the further one moves to the right, the harder it becomes to discern which box was looked at the most. The least informative signal possible corresponds to allocating the same amount of viewing time on each of the four boxes, i.e., 75% on all but the most inspected box. This point is marked in the figure by the dashed lines. The CDFs from the hide & seek games start out relatively low and with a flat slope which becomes steeper toward these dashed lines. This indicates that at least some first movers tried to blur their intended choice by looking at each of the four boxes for approximately the same amount of time.

Especially when the number of transitions between boxes increases, it becomes exceedingly more difficult to mentally keep track of which is the most inspected box. Among the first movers that look at all four boxes at least once, the mean number of fixations is in fact very high in both *LiveGaze* treatments (13.8 in *ForcedChoice* and 18.4 in *FreeChoice*), indicating that these first movers are indeed trying to make themselves unpredictable. For our further analyses, we therefore categorize first movers that looked at all four boxes at least once as signal-jammers.[29] The majority of first movers in both *LiveGaze* treatments apply this method (56.7% in *ForcedChoice* and 66.1% in *FreeChoice*, also see Table 5).

A second straight forward method to hide one's intended choice in our setting would be to simply not look at any boxes at all, e.g., by closing your eyes and blindly pressing a button. We did not mention this possibility in the instructions but we neither explicitly forbade it. Nevertheless, only 6.1% of first movers applied this method in the *LiveGaze-ForcedChoice* treatment and only 8.9% in the *LiveGaze-FreeChoice* treatment.

Apart from trying to hide their intentions, first movers could also try to actively mislead second movers into a wrong belief about their intended choice. To achieve this, some participants tried to emphasize one or more boxes with their gaze and then choose either the emphasized or a different box. In the post-experimental questionnaire, participants using such techniques for instance described them like this: "[I] never looked at the chosen box", "[I looked] at the chosen box because I assumed that the respective other partner would not expect this", "I concentrated on a different box and switched to another box shortly

[29] Only very few first movers looked at each of the four boxes only exactly once (2.8% and 1.1% of hide & seek trials in *LiveGaze-ForcedChoice* and *LiveGaze-FreeChoice*, respectively). Among the four first movers that played in these trials, two explicitly stated that they tried to look at each box for the same amount of time in the post-experimental questionnaire ("[I looked] at all boxes once, one after another, each equally long.", "I attempted to look at all boxes equally long such that nothing can be derived from it.") and one allocated his viewing time particularly evenly across the four boxes (standard deviation of box inspection times below 2% in all trials).

before my choice but also did not choose that one", "I always looked at a different box first and then briefly at the one I chose and then at the first one again". We therefore consider all those first movers as active misleaders who only looked at one, two, or three boxes in the hide & seek game. In the *LiveGaze-ForcedChoice* treatment, 20% of first movers only looked at a single box and 17.2% looked at two or three boxes. Similarly, in the *LiveGaze-FreeChoice* treatment, 13.9% looked at one box and 11.1% looked at two or three. Table 5 summarizes which strategies have been applied in the two *LiveGaze* treatments and also shows their respective success rates in parentheses.

TABLE 5—FIRST MOVERS' GAZE STRATEGIES AND SUCCESS RATES (IN PARENTHESES) IN HIDE & SEEK GAME.

LiveGaze-ForcedChoice, First Movers					
	Applying	Share of which Choosing:			
Gaze Strategy	Strategy	Most = Last	Most Not Last	Last Not Most	Neither
Did Not Look	6.1% (63.6%)				
Looked At All	56.7% (79.4%)	9.8% (50.0%)	1.0% (100%)	14.7% (73.3%)	74.5% (84.2%)
Looked At One	20.0% (63.9%)	44.4% (31.3%)	-	-	55.6% (90.0%)
Other	17.2% (77.4%)	29.0% (77.8%)	3.2% (0%)	6.5% (50.0%)	61.3% (84.2%)
Mean	100% (75.0%)	19.4% (48.6%)	1.1% (50.0%)	9.4% (70.6%)	70.0% (83.3%)

LiveGaze-FreeChoice, First Movers					
	Applying	Share of which Choosing:			
Gaze Strategy	Strategy	Most = Last	Most Not Last	Last Not Most	Neither
Did Not Look	8.9% (75.0%)				
Looked At All	66.1% (68.9%)	16.0% (63.2%)	5.9% (57.1%)	10.9% (61.5%)	67.2% (72.5%)
Looked At One	13.9% (76.0%)	48.0% (75.0%)	-	-	52.0% (76.9%)
Other	11.1% (70.0%)	20.0% (75.0%)	0%	10.0% (100%)	70.0% (64.3%)
Mean	100% (70.6%)	19.4% (68.6%)	3.9% (57.1%)	8.3% (66.7%)	68.3% (72.4%)

Notes: The first column shows the share of first movers applying the respective gaze strategy. The next four columns show the box choices conditional on the gaze strategy: "Most = Last" is the share of first movers choosing the box they looked at for the longest time if this is also the last one they looked at, "Most Not Last" is the share choosing the box they looked at most if this doesn't coincide with the last one, "Last Not Most" is the share choosing the box they looked at last if this doesn't coincide with the most inspected one, and "Neither" is the share choosing neither of the two. The numbers in parentheses correspond to the first movers' success rates conditional on their gaze strategy (first column) and choice (other columns).

Recall that, on average, first movers in both *LiveGaze* treatments were able to reach success rates statistically indistinguishable from those in the *NoGaze* treatment (70.5%). However, in the *LiveGaze-ForcedChoice* treatment, much higher success rates were reached by first movers who chose neither the box they inspected most nor last. Subjects that looked at only one box, but then chose a different one even reached an average success rate of 90%. As can be seen in Table 6, this effect is also significant. It shows the marginal effects of several probit models of first movers' success in the hide & seek game regressed on the different strategies. The baseline choice strategy that is left out in the models is choosing the most emphasized box (i.e., the box that was looked at last when this was also the box that was looked at most).

TABLE 6—FIRST MOVER SUCCESS REGRESSED ON GAZE STRATEGY (PROBIT MODELS, MARGINAL EFFECTS).

LiveGaze-ForcedChoice				
	Probit Dependent Variable: First Mover Success (binary)			
(Sub-)Sample:	Full	Looked at All	Looked at One	Other
Most Not Last	.010	(omitted)†	-	(omitted)†
	(.087)			
Last Not Most	.140*	.117*	-	-.174
	(.083)	(.069)		(.435)
Neither	.337***	.283**	.588***	.145
	(.121)	(.128)	(.154)	(.201)
N	180	102	36	31

LiveGaze-FreeChoice				
	Probit Dependent Variable: First Mover Success (binary)			
(Sub-)Sample:	Full	Looked at All	Looked at One	Other
Most Not Last	-.104	-.055	-	(omitted)†
	(.176)	(.197)		
Last Not Most	-.018	-.015	-	(omitted)†
	(.153)	(.173)		
Neither	.038	.091	.019	-.111
	(.083)	(.124)	(.168)	(.314)
N	180	119	25	18

Notes: Baseline category is "Most = Last", i.e., choosing the box that was looked at most when it coincides with the box that was looked at last. Explanatory variables are dummies corresponding to the other choice strategies. The same model is estimated on different sub-samples. The whole sample was used in the models in the first column. In the second column, only those trials were included in which all four boxes were inspected at least once. In the third column, only those in which exactly one box was looked at and in the last column all trials in which two or three boxes were inspected. Robust standard errors clustered at the subject level. Standard errors for marginal effects in parentheses obtained via the Delta-Method.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.
†Omitted due to too few observations in cell.

The top half of the table shows the results for the *LiveGaze-ForcedChoice* treatment. From the first column we can see that the probability for a first mover success is significantly higher by more than 30% if he chooses neither the last nor the most inspected box. When we look at the same model estimated for different sub-samples, we see that this effect is especially pronounced for those subjects that only looked at one single box but then chose a different one. These subjects even significantly increased their success probability by 58.8% compared to those that chose the only box they looked at (column "Looked at One"). The effect of choosing neither the most nor the last inspected box is also significantly positive for those subjects who looked at all four boxes at least once

(+28.3%, column “Looked at All”). For the same group of subjects, also choosing the box they looked at last significantly increased their success probability by 11.7% when this box did *not* coincide with their most inspected box. This indicates that not all first movers that tried to blur their intention by looking at all four boxes succeeded in doing so and that the second movers still managed to deduce some meaning from their gaze.[30]

This does not seem to be the case in the *LiveGaze-FreeChoice* treatment, however. Here, none of the strategies significantly affected the first movers’ success probabilities, neither in the whole sample nor in any of the sub-samples. At least the tendencies seem to be the same though: Both when only one or when all of the boxes were looked at, success rates are highest when neither the most nor the last inspected box was chosen. It thus seems that even though we do find evidence that some people try to use their gaze in order to actively mislead their counterpart, only few succeed and only in specific circumstances. More precisely, the ability in the *LiveGaze-ForcedChoice* treatment to fully focus on one’s gaze strategy without having to worry about the actual choice (which is practically made by the computer) seems to help at least some subjects in choosing a good gaze strategy.

V. Conclusion

We have shown that people are generally able to correctly interpret non-strategic gaze in common interest games and, to a lesser extent, in a competitive hide & seek game. The differences in the degree to which non-strategic gaze is understood can be explained by differences in the non-strategic gaze patterns: Subjects were more predictable when they chose the option that they looked at last or for the longest time.

---

[30] Indeed, still more than 50% of second movers chose the box that the first mover either looked at most or last (or both) in the hide & seek game in the *LiveGaze-ForcedChoice* treatment even when the first movers tried to blur their intention and looked at all four boxes (see Table 11 in Appendix C).

Subjects strategically adapted their eye movements when they were live transmitted to their current opponent. In the common interest games, where it is in their interest to be predictable, they try to make their gaze particularly easy to interpret, e.g., by lowering the number of fixations while increasing fixation durations or by looking at just one single option. In the hide & seek game, most subjects try to hide their intended choice from their eye movements by doing the exact opposite or by not looking at any option at all. Fewer subjects tried to actively mislead their opponents in the hide & seek game by sending a clear signal in order to induce a wrong belief.

Compared to non-strategic gaze, strategic gazing led to even higher success rates in the coordination game. Additionally, the inefficient focal outcome was avoided almost entirely in the focal coordination game. In the focal discoordination game, no higher success rates could be achieved than with non-strategic gaze. Again, the differences can be explained by differences in gaze patterns as subjects exhibit more ambiguous eye movements in the discoordination game than in the coordination games.

In the hide & seek game, the strategically adapted eye movements led to making the first movers less predictable. Their average success rate was statistically indistinguishable from what would be expected under random play. However, at least in the *LiveGaze-ForcedChoice* treatment, participants significantly increased their chances of winning by not choosing an option they emphasized with their gaze. This is not the case when subjects had to choose an option without the help of the computer, though. More research is needed in order to investigate whether strategically deceptive signals (be it with the eyes or via any other cues) generally impede the quality of the subsequent decisions. Since some of the gaze strategies were in fact successful, it would also be interesting to test whether applying (and dodging) such strategies can be trained. This could be especially relevant for professional sports with many one-on-one situations in which being predictable is particularly costly.

Another open question is whether gaze can also signal things other than intentions such as sophistication, commitment, trustworthiness, attitude or cooperativeness: Can people discern whether or not the eye movements they see stem from an expert? Or a person who is particularly devoted to a task? Are they able to predict how trustworthy or cooperative a person will act judging from her gaze? It would then again be interesting to see whether live gaze transmission is strategically utilized by the eye-tracked subjects also in these settings, e.g., to signal their own or a different type in order to evoke a specific response.

Appendix

A. Nash Equilibria

TABLE 7—COORDINATION GAME, NASH EQUILIBRIA.

Player 1				Player 2			
TL	TR	BL	BR	TL	TR	BL	BR
1/4	1/4	1/4	1/4	1/4	1/4	1/4	1/4
0	1/3	1/3	1/3	0	1/3	1/3	1/3
1/3	0	1/3	1/3	1/3	0	1/3	1/3
1/3	1/3	0	1/3	1/3	1/3	0	1/3
1/3	1/3	1/3	0	1/3	1/3	1/3	0
0	0	1/2	1/2	0	0	1/2	1/2
0	1/2	0	1/2	0	1/2	0	1/2
1/2	0	0	1/2	1/2	0	0	1/2
0	1/2	1/2	0	0	1/2	1/2	0
1/2	0	1/2	0	1/2	0	1/2	0
1/2	1/2	0	0	1/2	1/2	0	0
1	0	0	0	1	0	0	0
0	1	0	0	0	1	0	0
0	0	1	0	0	0	1	0
0	0	0	1	0	0	0	1

Notes: Each row corresponds to one equilibrium. The numbers show the support for the respective strategy in the respective equilibrium.

TABLE 8—FOCAL COORDINATION GAME, NASH EQUILIBRIA.

Player 1				Player 2			
TL	TR	BL	BR	TL	TR	BL	BR
11/41	10/41	10/41	10/41	11/41	10/41	10/41	10/41
0	1/3	1/3	1/3	0	1/3	1/3	1/3
11/31	0	10/31	10/31	11/31	0	10/31	10/31
11/31	10/31	0	10/31	11/31	10/31	0	10/31
11/31	10/31	10/31	0	11/31	10/31	10/31	0
0	0	1/2	1/2	0	0	1/2	1/2
0	1/2	0	1/2	0	1/2	0	1/2
11/21	0	0	10/21	11/21	0	0	10/21
0	1/2	1/2	0	0	1/2	1/2	0
11/21	0	10/21	0	11/21	0	10/21	0
11/21	10/21	0	0	11/21	10/21	0	0
1	0	0	0	1	0	0	0
0	1	0	0	0	1	0	0
0	0	1	0	0	0	1	0
0	0	0	1	0	0	0	1

Notes: Each row corresponds to one equilibrium. The numbers show the support for the respective strategy in the respective equilibrium. TL was the focal option.

TABLE 9—HIDE & SEEK GAME, NASH EQUILIBRIA.

Player 1				Player 2			
TL	TR	BL	BR	TL	TR	BL	BR
1/4	1/4	1/4	1/4	1/4	1/4	1/4	1/4

Notes: The hide & seek game only has one unique equilibrium in which both players play each strategy with equal probabilities.

TABLE 10—FOCAL DISCOORDINATION GAME, NASH EQUILIBRIA.

Player 1				Player 2			
TL	TR	BL	BR	TL	TR	BL	BR
8/41	11/41	11/41	11/41	8/41	11/41	11/41	11/41
1	0	0	0	0	9/11	1/11	1/11
1	0	0	0	0	1/11	9/11	1/11
1	0	0	0	0	1/11	1/11	9/11
0	9/11	1/11	1/11	1	0	0	0
0	1/11	9/11	1/11	1	0	0	0
0	1/11	1/11	9/11	1	0	0	0
0	1	0	0	0	0	1	0
0	1	0	0	0	0	0	1
0	0	1	0	0	1	0	0
0	0	1	0	0	0	0	1
0	0	0	1	0	1	0	0
0	0	0	1	0	0	1	0

Notes: Each row corresponds to one equilibrium. The numbers show the support for the respective strategy in the respective equilibrium. TL was the focal option.

There are four Nash Equilibria in pure strategies in the coordination game but strategies are isomorphic with no reason to choose one over the others except possibly the labeling of the strategies, which here corresponds to the boxes' locations. Similar arguments apply to the focal games but here the focal outcome (TL) could be used as a coordination device.

The hide & seek game is asymmetric. This game has a unique Nash equilibrium in mixed strategies with both players choosing all strategies with equal probability. Since players play this game without feedback, all strategy choices are consistent with equilibrium. Nevertheless, if all players play according to the Nash equilibrium, one would expect that strategy choices are approximately uniformly distributed in the aggregate.

B. Detailed Procedures and Instructions

Before entering the lab, all participants were welcomed by the experimenters and given oral instructions about the general rules and procedures inside the lab. If the session included eye-tracking, the experimenters also showed the participants one of the chin rests and introduced them to the general eye-tracking procedures. After an ID-check to assure that all participants were properly registered to the experiment, they were randomly allocated to individual booths. At their booths they found brief written instructions reminding them of the general rules in the lab and announcing that the experiment will comprise four parts and one questionnaire in the end. This questionnaire was merely used to keep participants busy while the experimenters prepared the receipts. Subjects that were randomly allocated to one of the booths with an eye-tracker further received written instructions regarding the calibration procedure. Once the participants were finished reading these instructions, participants in the *LiveGaze-FreeChoice* treatment received the following oral instructions read out by one of the experimenters (instructions for the other treatments as well as all written instructions are available upon request):

> *Before we begin we will briefly explain the structure of the experiment. Today there are two types of participants: A and B. The first half—who have an eye-tracker—are participants A and the other half are participants B.*
>
> *There will be 4 parts, each with 5 rounds and in each round, one participant A and one participant B are in a group together. The composition of the groups is randomly redrawn before every round.*
>
> *All rounds have the same general structure—only the income changes. At the end of the experiment you receive the points you earned in all rounds in cash. 10 points are worth 1 Euro.*
>
> *You will learn how much you earned at the end of the experiment. This means there will be no feedback between rounds.*
>
> *When you look at your screens now, you will see four gray boxes. In each round, both participants sequentially choose one of the four boxes: First participant A and then participant B.*
>
> *On each box you see the amount of points you could theoretically earn when you choose this box. You will always either earn the amount displayed on the box you choose or zero points. The numbers on the boxes are always the same for both participants. So if, for instance, you see a 10 in the top left box, then there will always also be a 10 in the top left box for the other participant.*
>
> *Before each of the four parts you will learn how exactly your income will be determined. In general, it will depend on whether participants A and B will choose the same or a different box.*
>
> *There is a special feature for participants A: As soon as the eye-trackers are running, the box that they are currently looking at will be red instead of gray.*

Also participant B can see which box participant A is looking at until participant A has chosen a box. This means that participant B always sees the same screen or rather the same red boxes flashing as participant A.

For example, if participant A looks at the top right box, this box will turn red. If participant A then looks at another box, this one will turn red and the first one will turn gray again—and that is both for participant A and participant B. Hence, participant B knows which box participant A is currently looking at.

Participant B can only decide once participant A is done. This will be signaled to him by a green frame that appears around the four boxes. On participant B's screen this will look like this. [The green frame was then displayed on all participants' screens.]

As soon as this green frame appears, participant B can decide. To choose a box, all participants use the four keys with the stickers on their keyboards. To choose the top left box, you have to press the top left key, for the top right box, the top right key, and so forth.

We will now begin with the calibration of the eye-trackers. This will be followed by four practice rounds to familiarize you with the controls and procedures. These practice rounds have no influence on your income. We ask all participants B to wait while participants A are being calibrated.

The four trial rounds mentioned in the instructions existed in all treatments. Their only purpose was to make the participants accustomed to the input method and the look and feel of the visualization of the gaze data. For the first movers, their true gaze data was visualized already in these practice rounds, whereas second movers saw a randomly generated gaze pattern (except for the *NoGaze* treatment in which the second movers saw no gaze patterns but simply had to wait for an equivalent amount of time). Another difference between the practice rounds and the actual experiment was that in each of the practice rounds participants saw an arrow pointing at one of the four boxes. In order to advance to the next round, they had to press the key that corresponded to the box with the arrow.

The instructions concerning the individual games were shown on screen before each part and read like this (here, we show the instructions for the case that the coordination game was played in part 1):

Part 1 (rounds 1 to 5)

In Part 1, your income will be determined like this:
If participant B chooses the same box as you, you will earn 10 points.
If participant B chooses a different box than you, you will earn 0 points.

Participant B's income will be determined like this:
If participant B chooses the same box as you, he will earn 10 points.
If participant B chooses a different box than you, he will earn 0 points.

If you have any questions, please raise your hand, otherwise you can now press Space to continue.

Before each individual round, there was an additional reminder of the random re-matching of partners ("*You will now be randomly matched with a new participant A[B]*.") as well as of the income structure ("*Participant A earns points, if the chosen boxes are the same[different]. // Participant B earns points, if the chosen boxes are the same[different].*"). This screen was shown for eight seconds, followed by a screen announcing the beginning of the next round for two seconds, and finally the next round started with the display of a fixation cross in the center of the screen for one second and then the four boxes were displayed.

C. Further Analyses

TABLE 11—SHARE OF SECOND MOVER CHOICES IN LINE WITH HEURISTICS BY TREATMENT.

	RecordedGaze			
	Coordination	Focal Coordination	Focal Discoord.	Hide & Seek
Most Looked	0.81	0.77	0.93	0.68
Last Looked	0.86	0.82	0.94	0.67
Frugal Tree	0.86	0.80	0.92	0.66

	LiveGaze-ForcedChoice			
	Coordination	Focal Coordination	Focal Discoord.	Hide & Seek
Most Looked	0.94	0.96	0.93	0.46
Last Looked	0.94	0.94	0.91	0.45
Frugal Tree	0.94	0.94	0.92	0.45

	LiveGaze-FreeChoice			
	Coordination	Focal Coordination	Focal Discoord.	Hide & Seek
Most Looked	0.96	0.99	0.93	0.31
Last Looked	0.94	0.99	0.91	0.30
Frugal Tree	0.96	0.99	0.92	0.33

Notes: A second mover's choice is counted as in line with the heuristic, if it is a best response to the heuristic's predicted choice of the first mover. Note that, since the heuristics always make point predictions for the first mover, there always are three best responses in the focal discoordination game whereas there is only one in all other games.

The share of second mover decisions in line with the respective heuristics never differ significantly between the coordination and the focal coordination game in none of the treatments (all corrected p-values > .1). In both *LiveGaze* treatments, all three heuristics' hit rates with respect to second mover behavior are significantly lower in the hide & seek game than in all other games (all corrected p-values < .002). Additionally, in the *LiveGaze-FreeChoice* treatment, hit rates of all three heuristics are also significantly higher in the focal coordination game than in the focal discoordination game (all corrected p-values < .05).

In the *RecordedGaze* treatment, the hit rate of the *Most Looked* heuristic is significantly higher in the focal discoordination game than in the other three games (all corrected p-values < .05). Similarly, also in the *RecordedGaze* treatment, the hit rates of the *Last Looked* heuristic and the *Frugal Tree* are significantly higher in the focal discoordination game than in the focal coordination game (both corrected p-values < .05) but not compared to the coordination game (both corrected p-values > .6). Further, the hit rate of the *Last Looked* heuristic is significantly lower in the hide & seek game than in all other games (all corrected p-values < .1). Finally, the *Frugal Tree*'s hit rate is significantly lower in the *RecordedGaze* hide & seek game than in the focal discoordination and the coordination game (both corrected p-values < .02) but not compared to its hit rate in the focal coordination game (corrected p-value > .1). All unmentioned comparisons are insignificant.

References

Abed, F. (1991). Cultural influences on visual scanning patterns. *Journal of Cross-Cultural Psychology, 22*(4), 525-534.

Bardsley, N., Mehta, J., Starmer, C., & Sugden, R. (2010). Explaining focal points: Cognitive hierarchy theory versus team reasoning. *The Economic Journal, 120*(543), 40-79.

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition, 106*(3), 1465-1477.

Brocas, I., Carrillo, J. D., Wang, S. W., & Camerer, C. F. (2014). Imperfect choice or imperfect attention? Understanding strategic thinking in private information games. *Review of Economic Studies, 81*(3), 944-970.

Cañal-Bruland, R., & Schmidt, M. (2009). Response bias in judging deceptive movements. *Acta Psychologica, 130*(3), 235-240.

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences, 102*(35), 12629-12633.

Costa-Gomes, M., Crawford, V. P., & Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica, 69*(5), 1193-1235.

Crawford, V. P., & Iriberri, N. (2007). Fatal attraction: Salience, naïveté, and sophistication in experimental “hide-and-seek” games. *American Economic Review, 97*(5), 1731-1750.

Dicks, M., Button, C., & Davids, K. (2010). Availability of advance visual information constrains association-football goalkeeping performance during penalty kicks. *Perception, 39*(8), 1111-1124.

Emara, N., Owens, D., Smith, J., & Wilmer, L. (2017). Serial correlation in National Football League play calling and its effects on outcomes. *Journal of Behavioral and Experimental Economics, 69*, 125-132.

Farrell, J., & Rabin, M. (1996). Cheap talk. *Journal of Economic Perspectives, 10*(3), 103-118.

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

Foulsham, T., & Lock, M. (2015). How the eyes tell lies: Social gaze during a preference task. *Cognitive Science, 39*(7), 1704-1726.

Frydman, C., & Krajbich, I. (2017). Using response times to infer others' beliefs: An application to information cascades. *Mimeo*.

Gallagher-Mitchell, T., Simms, V., & Litchfield, D. (2017). Learning from where 'eye' remotely look or point: Impact on number line estimation error in adults. *Quarterly Journal of Experimental Psychology,* (online first).

Greiner, B. (2004). The online recruitment system orsee 2.0–A guide for the organization of experiments in economics. *University of Cologne, Working Paper Series in Economics, 10*(23), 63-104.

Holmes, K. J., Ayzenberg, V., & Lourenco, S. F. (2016). Gamble on gaze: Eye movements reflect the numerical value of blackjack hands. *Psychonomic Bulletin & Review, 23*(6), 1974-1981.

Ishii, Y., Okubo, M., Nicholls, M. E., & Imai, H. (2011). Lateral biases and reading direction: A dissociation between aesthetic preference and line bisection. *Brain and Cognition, 75*(3), 242-247.

Isoni, A., Poulsen, A., Sugden, R., & Tsutsui, K. (2013). Focal points in tacit bargaining problems: Experimental evidence. *European Economic Review, 59*, 167-188.

Jackson, R. C., Warren, S., & Abernethy, B. (2006). Anticipation skill and susceptibility to deceptive movement. *Acta Psychologica, 123*(3), 355-371.

Knoepfle, D. T., Wang, J. T.-Y., & Camerer, C. F. (2009). Studying learning in games using eye-tracking. *Journal of the European Economic Association, 7*(2-3), 388-398.

Kovács-Bálint, Z., Bereczkei, T., & Hernádi, I. (2013). The telltale face: Possible mechanisms behind defector and cooperator recognition revealed by emotional facial expression metrics. *British Journal of Psychology, 104*(4), 563-576.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience, 13*(10), 1292-1298.

Lahey, J. N., & Oxley, D. (2016). The power of eye tracking in economics experiments. *American Economic Review, 106*(5), 309-313.

Levitt, S. D., List, J. A., & Reiley, D. H. (2010). What happens in the field stays in the field: Exploring whether professionals play minimax in laboratory experiments. *Econometrica, 78*(4), 1413-1434.

Litchfield, D., & Ball, L. J. (2011). Using another's gaze as an explicit aid to insight problem solving. *Quarterly Journal of Experimental Psychology, 64*(4), 649-656.

Litchfield, D., Ball, L. J., Donovan, T., Manning, D. J., & Crawford, T. (2010). Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *Journal of Experimental Psychology: Applied, 16*(3), 251-262.

Liu, C. (2008). *The Dark Forest* (J. Martinsen, Trans.). New York: Tor (translated version, 2015).

Müller, R., Helmert, J. R., & Pannasch, S. (2014). Limitations of gaze transfer: Without visual context, eye movements do not help to coordinate joint action, whereas mouse movements do. *Acta Psychologica, 152*, 19-28.

Müller, R., Helmert, J. R., Pannasch, S., & Velichkovsky, B. M. (2013). Gaze transfer in remote cooperation: Is it always helpful to see what your partner is attending to? *Quarterly Journal of Experimental Psychology, 66*(7), 1302-1316.

Nagano, T., Kato, T., & Fukuda, T. (2006). Visual behaviors of soccer players while kicking with the inside of the foot. *Perceptual and Motor Skills, 102*(1), 147-156.

Nalanagula, D., Greenstein, J. S., & Gramopadhye, A. K. (2006). Evaluation of the effect of feedforward training displays of search strategy on visual search performance. *International Journal of Industrial Ergonomics, 36*(4), 289-300.

Neider, M. B., Chen, X., Dickinson, C. A., Brennan, S. E., & Zelinsky, G. J. (2010). Coordinating spatial referencing using shared gaze. *Psychonomic Bulletin & Review, 17*(5), 718-724.

Penczynski, S. P. (2016). Strategic thinking: The influence of the game. *Journal of Economic Behavior & Organization, 128*, 72-84.

Polonio, L., Di Guida, S., & Coricelli, G. (2015). Strategic sophistication and attention in games: An eye-tracking study. *Games and Economic Behavior, 94*, 80-96.

Raab, M., & Gigerenzer, G. (2015). The power of simplicity: A fast-and-frugal heuristics approach to performance science. *Frontiers in Psychology, 6*(1672), 1-6.

Rapoport, A., & Budescu, D. V. (1992). Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General, 121*(3), 352-363.

Rowe, R., Horswill, M. S., Kronvall-Parkinson, M., Poulter, D. R., & McKenna, F. P. (2009). The effect of disguise on novice and expert tennis players' anticipation ability. *Journal of Applied Sport Psychology, 21*(2), 178-185.

Rubinstein, A., Tversky, A., & Heller, D. (1997). Naive strategies in competitive games. In W. Albers, W. Güth, P. Hammerstein, B. Moldovanu, & E. van Damme (Eds.), *Understanding Strategic Interaction: Essays in Honor of Reinhard Selten* (pp. 394-402). Berlin, Heidelberg: Springer.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, 71-78.

Schelling, T. C. (1960). *The Strategy of Conflict*. Cambridge, MA; London, England: Harvard University Press (reprinted and edited version, 1980).

Schwedes, C., & Wentura, D. (2012). The revealing glance: Eye gaze behavior to concealed information. *Memory & Cognition, 40*(4), 642-651.

Sebanz, N., & Shiffrar, M. (2009). Detecting deception in a bluffing body: The role of expertise. *Psychonomic Bulletin & Review, 16*(1), 170-175.

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience, 6*(12), 1317-1322.

Stein, R., & Brennan, S. E. (2004). Another person's eye gaze as a cue in solving programming problems. *Proceedings of the 6th International Conference on Multimodal Interface*, 1-7.

Stewart, N., Gächter, S., Noguchi, T., & Mullett, T. L. (2016). Eye movements in strategic choice. *Journal of Behavioral Decision Making, 29*(2-3), 137-156.

Tay, C. S., Chow, J. Y., Koh, M., & Button, C. (2010). The effectiveness of keeper-independent penalty kicks using fake visual cues from penalty takers. *International Journal of Sport Psychology, 43*(5), 403-419.

Van Leeuwen, B., Noussair, C. N., Offerman, T., Suetens, S., van Veelen, M., & van de Ven, J. (2017). Predictably angry–Facial cues provide a credible signal of destructive behavior. *Management Science,* (online first).

Velichkovsky, B. M. (1995). Communicating attention: Gaze position transfer in cooperative problem solving. *Pragmatics & Cognition, 3*(2), 199-223.

Vrij, A. (2004). Why professionals fail to catch liars and how they can improve. *Legal and Criminological Psychology, 9*(2), 159-181.

Walker, M., & Wooders, J. (2001). Minimax play at Wimbledon. *American Economic Review, 91*(5), 1521-1538.

Wang, J. T.-Y., Spezio, M., & Camerer, C. F. (2010). Pinocchio's pupil: Using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games. *American Economic Review, 100*(3), 984-1007.

Wermeskerken, M., Litchfield, D., & Gog, T. (2017). What Am I Looking at? Interpreting Dynamic and Static Gaze Displays. *Cognitive Science,* (online first).

Abgrenzung

Der erste Artikel „Processes and Preferences Underlying Third Party Rewards and Punishments” enstammt einer Zusammenarbeit mit Urs Fischbacher. Dabei wurde das experimentelle Design gemeinsam entworfen. Die Instruktionen wurden von mir geschrieben und das Experiment wurde von mir programmiert und durchgeführt, alles in enger Absprache mit Urs Fischbacher. Die Datenanalyse wurde von mir durchgeführt und regelmäßig mit Urs Fischbacher besprochen. Der Artikel wurde größtenteils von mir verfasst, wir haben ihn jedoch noch oft diskutiert und ich habe ihn auf dieser Grundlage überarbeitet.

Der zweite Artikel „In Vino Veritas“ ist in Kollaboration mit Jan Hausfeld entstanden. Dabei wurde das experimentelle Design gemeinsam entworfen. Die Programmierung der Tablets und die Instruktionen stammen von mir und die Experimente wurden von uns gemeinsam durchgeführt. Die Zusammenführung der Daten haben wir gemeinsam durchgeführt und die anschließende Datenanalyse stammt von mir. Das Papier wurde von mir geschrieben und abschließend mit Jan Hausfeld überarbeitet.

Der dritte Artikel „Strategic Gaze: An Interactive Eye-Tracking Study“ wurde in Zusammenarbeit mit Jan Hausfeld und Susanne Goldlücke verfasst. Das experimentelle Design haben wir zu dritt erarbeitet. Die Programmierung und das Verfassen der Instruktionen habe ich übernommen. Die Experimente wurden von mir und Jan Hausfeld gemeinsam durchgeführt. Die Zusammenführung der Daten haben Jan Hausfeld und ich gemeinsam erledigt. Anschließend haben wir beide die Daten unabhängig voneinander analysiert. Das Papier wurde dann von uns gemeinsam geschrieben, wobei Kapitel 2 bis 5 größtenteils aus meiner Hand stammen.

Bibliography

Abed, F. (1991). Cultural influences on visual scanning patterns. *Journal of Cross-Cultural Psychology, 22*(4), 525-534.

Abeler, J., Nosenzo, D., & Raymond, C. (2016). Preferences for truth-telling. *CESifo Working Paper Series, No. 6087*, 1-118.

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters, 80*(1), 123-129.

Aldrich, J. H., & Nelson, F. D. (1984). *Linear Probability, Logit, and Probit Models* (Vol. 45). Newbury Park, London, New Delhi: Sage.

Almenberg, J., Dreber, A., Apicella, C., & Rand, D. G. (2011). Third party reward and punishment: group size, efficiency and public goods. In N. M. Palmetti & J. P. Russo (Eds.), *Psychology of Punishment* (pp. 73-92): Nova Science Publishers.

Anderson, L. R., & Mellor, J. M. (2008). Predicting health behaviors with an experimental measure of risk preference. *Journal of Health Economics, 27*(5), 1260-1274.

Andreoni, J., Harbaugh, W. T., & Vesterlund, L. (2003). The carrot or the stick: Rewards, punishments and cooperation. *American Economic Review, 93*(3), 893-902.

Arbel, Y., Bar-El, R., Siniver, E., & Tobol, Y. (2014). Roll a die and tell a lie–What affects honesty? *Journal of Economic Behavior & Organization, 107*, 153-172.

Arthur, W., & Day, D. V. (1994). Development of a short form for the Raven Advanced Progressive Matrices Test. *Educational and Psychological Measurement, 54*(2), 394-403.

Bardsley, N., Mehta, J., Starmer, C., & Sugden, R. (2010). Explaining focal points: Cognitive hierarchy theory versus team reasoning. *The Economic Journal, 120*(543), 40-79.

Barfort, S., Harmon, N., Hjorth, F., & Olsen, A. L. (2015). Dishonesty and selection into public service in Denmark: Who runs the world's least corrupt public sector? *University of Copenhagen Discussion Paper Series, 15*(12), 1-56.

Barsky, R. B., Juster, F. T., Kimball, M. S., & Shapiro, M. D. (1997). Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. *Quarterly Journal of Economics, 112*(2), 537-579.

Bartling, B., Engl, F., & Weber, R. A. (2014). Does willful ignorance deflect punishment?–An experimental study. *European Economic Review, 70*, 512-524.

Bègue, L., Bushman, B. J., Zerhouni, O., Subra, B., & Ourabah, M. (2013). 'Beauty is in the eye of the beer holder': People who think they are drunk also think they are attractive. *British Journal of Psychology, 104*(2), 225-234.

Bendor, J., & Swistak, P. (2001). The evolution of norms. *American Journal of Sociology, 106*(6), 1493-1545.

Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. *Nature, 442*, 912-915.

Biziou-van-Pol, L., Haenen, J., Novaro, A., Liberman, A. O., & Capraro, V. (2015). Does telling white lies signal pro-social preferences? *Judgment and Decision Making, 10*(6), 538-548.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review, 90*(1), 166-193.

Bregu, K., Deck, C., Ham, L., & Jahedi, S. (2017). The effects of alcohol use on economic decision making. *Southern Economic Journal, 83*(4), 886-902.

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition, 106*(3), 1465-1477.

Breslin, F. C., Sobell, M. B., Cappell, H., Vakili, S., & Poulos, C. X. (1999). The effects of alcohol, gender, and sensation seeking on the gambling choices of social drinkers. *Psychology of Addictive Behaviors, 13*(3), 243-252.

Brocas, I., Carrillo, J. D., Wang, S. W., & Camerer, C. F. (2014). Imperfect choice or imperfect attention? Understanding strategic thinking in private information games. *Review of Economic Studies, 81*(3), 944-970.

Buckholtz, J. W., & Marois, R. (2012). The roots of modern justice: Cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience, 15*(5), 655-661.

Burghart, D. R., Glimcher, P. W., & Lazzaro, S. C. (2013). An expected utility maximizer walks into a bar. *Journal of Risk and Uncertainty, 46*(3), 215-246.

Butler, J. V., Conzo, P., & Leroch, M. (2015). Social Identity and Punishment. *GSME Discussion Paper Series, No. 1512*.

Cañal-Bruland, R., & Schmidt, M. (2009). Response bias in judging deceptive movements. *Acta Psychologica, 130*(3), 235-240.

Cappelen, A. W., Sørensen, E. Ø., & Tungodden, B. (2013). When do we lie? *Journal of Economic Behavior & Organization, 93*, 258-265.

Capraro, V. (2017a). Does the truth come naturally? Time pressure increases honesty in one-shot deception games. *Economics Letters*, *158*, 54-57.

Capraro, V. (2017b). Who lies? A meta-analysis of the effect of sex, age, and education on honesty. *Mimeo*.

Charness, G., Cobo-Reyes, R., & Jiménez, N. (2008). An investment game with third-party intervention. *Journal of Economic Behavior & Organization, 68*(1), 18-28.

Charness, G., & Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, *74*(6), 1579-1601.

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

Childs, J. (2012). Gender differences in lying. *Economics Letters, 114*(2), 147-149.

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences, 102*(35), 12629-12633.

Cohen, T. R., Gunia, B. C., Kim-Jun, S. Y., & Murnighan, J. K. (2009). Do groups lie more than individuals? Honesty and deception as a function of strategic self-interest. *Journal of Experimental Social Psychology, 45*(6), 1321-1324.

Cohn, A., & Maréchal, M. A. (2017). Laboratory measure of cheating predicts school misconduct. *Mimeo*.

Cohn, A., Maréchal, M. A., & Noll, T. (2015). Bad boys: How criminal identity salience affects rule violation. *Review of Economic Studies, 82*(4), 1289-1308.

Conrads, J., Irlenbusch, B., Rilke, R. M., & Walkowitz, G. (2013). Lying and team incentives. *Journal of Economic Psychology, 34*, 1-7.

Corazzini, L., Filippin, A., & Vanin, P. (2014). Economic behavior under alcohol influence: An experiment on time, risk, and social preferences. *IZA Discussion Paper Series, No. 8170*.

Costa-Gomes, M., Crawford, V. P., & Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica, 69*(5), 1193-1235.

Crawford, V. P., & Iriberri, N. (2007). Fatal attraction: Salience, naïveté, and sophistication in experimental "hide-and-seek" games. *American Economic Review, 97*(5), 1731-1750.

Curtin, J. J., & Fairchild, B. A. (2003). Alcohol and cognitive control: Implications for regulation of behavior during response conflict. *Journal of Abnormal Psychology, 112*(3), 424-436.

Dai, Z., Galeotti, F., & Villeval, M. C. (2017). Cheating in the lab predicts fraud in the field: An experiment in public transportation. *Management Science*, (forthcoming).

de Kwaadsteniet, E. W., Rijkhoff, S. A., & van Dijk, E. (2013). Equality as a benchmark for third-party punishment and reward: The moderating role of uncertainty in social dilemmas. *Organizational Behavior and Human Decision Processes, 120*(2), 251-259.

Denton, K., & Krebs, D. (1990). From the scene to the crime: The effect of alcohol and social context on moral judgment. *Journal of Personality and Social Psychology, 59*(2), 242-248.

Dicks, M., Button, C., & Davids, K. (2010). Availability of advance visual information constrains association-football goalkeeping performance during penalty kicks. *Perception, 39*(8), 1111-1124.

Dreber, A., & Johannesson, M. (2008). Gender differences in deception. *Economics Letters, 99*(1), 197-199.

Duke, A. A., & Bègue, L. (2015). The drunk utilitarian: Blood alcohol concentration predicts utilitarian responses in moral dilemmas. *Cognition, 134*, 121-127.

Emara, N., Owens, D., Smith, J., & Wilmer, L. (2017). Serial correlation in National Football League play calling and its effects on outcomes. *Journal of Behavioral and Experimental Economics, 69*, 125-132.

Erat, S., & Gneezy, U. (2012). White lies. *Management Science, 58*(4), 723-733.

Farrell, J., & Rabin, M. (1996). Cheap talk. *Journal of Economic Perspectives, 10*(3), 103-118.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*(2), 63-87.

Fehr, E., & Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives, 14*(3), 159-181.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*, 137-140.

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

Felson, R., Savolainen, J., Aaltonen, M., & Moustgaard, H. (2008). Is the association between alcohol use and delinquency causal or spurious? *Criminology, 46*(3), 785-808.

Fillmore, M. T., Carscadden, J. L., & Vogel-Sprott, M. (1998). Alcohol, cognitive impairment and expectancies. *Journal of Studies on Alcohol, 59*(2), 174-179.

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

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise–An experimental study on cheating. *Journal of the European Economic Association, 11*(3), 525-547.

Fligner, M. A., & Policello, G. E. (1981). Robust rank procedures for the Behrens-Fisher problem. *Journal of the American Statistical Association, 76*(373), 162-168.

Foerster, A., Pfister, R., Schmidts, C., Dignath, D., & Kunde, W. (2013). Honesty saves time (and justifications). *Frontiers in Psychology, 4*(473).

Fosgaard, T. R., Hansen, L. G., & Piovesan, M. (2013). Separating Will from Grace: An experiment on conformity and awareness in cheating. *Journal of Economic Behavior & Organization, 93*, 279-284.

Foulsham, T., & Lock, M. (2015). How the eyes tell lies: Social gaze during a preference task. *Cognitive Science, 39*(7), 1704-1726.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25-42.

Friesen, L., & Gangadharan, L. (2012). Individual level evidence of dishonesty and the gender effect. *Economics Letters, 117*(3), 624-626.

Frydman, C., & Krajbich, I. (2017). Using response times to infer others' beliefs: An application to information cascades. *Mimeo*.

Gächter, S., Renner, E., & Sefton, M. (2008). The long-run benefits of punishment. *Science, 322*, 1510.

Gächter, S., & Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature, 531*(7595), 496-499.

Galizzi, M. M., & Miraldo, M. (2017). Are you what you eat? Healthy behaviour and risk preferences. *The BE Journal of Economic Analysis & Policy, 17*(1).

Gallagher-Mitchell, T., Simms, V., & Litchfield, D. (2017). Learning from where 'eye' remotely look or point: Impact on number line estimation error in adults. *Quarterly Journal of Experimental Psychology,* (online first).

Gino, F., & Ariely, D. (2012). The dark side of creativity: Original thinkers can be more dishonest. *Journal of Personality and Social Psychology, 102*(3), 445-459.

Gino, F., Ayal, S., & Ariely, D. (2013). Self-serving altruism? The lure of unethical actions that benefit others. *Journal of Economic Behavior & Organization, 93*, 285-292.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review, 95*(1), 384-394.

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *Quarterly Journal of Economics, 112*(2), 631-645.

Goette, L., Huffman, D., & Meier, S. (2012). The impact of social ties on group interactions: Evidence from minimal groups and randomly assigned real groups. *American Economic Journal: Microeconomics, 4*(1), 101-115.

Greene, J. D., & Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences, 106*(30), 12506-12511.

Greiner, B. (2004). The online recruitment system orsee 2.0–A guide for the organization of experiments in economics. *University of Cologne, Working Paper Series in Economics, 10*(23), 63-104.

Gunia, B. C., Wang, L., Huang, L., Wang, J., & Murnighan, J. K. (2012). Contemplation and conversation: Subtle influences on moral decision making. *Academy of Management Journal, 55*(1), 13-33.

Gylfason, H. F., Arnardottir, A. A., & Kristinsson, K. (2013). More on gender differences in lying. *Economics Letters, 119*(1), 94-96.

Hanna, R., & Wang, S.-Y. (2013). Dishonesty and selection into public service. *NBER Working Paper Series, No. 19649*.

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., & Henrich, N. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science, 327*, 1480-1484.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., & Henrich, N. (2006). Costly punishment across human societies. *Science, 312*(5781), 1767-1770.

Hoff, K., Kshetramade, M., & Fehr, E. (2011). Caste and punishment: The legacy of caste culture in norm enforcement. *The Economic Journal, 121*(556), F449-F475.

Holmes, K. J., Ayzenberg, V., & Lourenco, S. F. (2016). Gamble on gaze: Eye movements reflect the numerical value of blackjack hands. *Psychonomic Bulletin & Review*, *23*(6), 1974-1981.

Hopthrow, T., Abrams, D., Frings, D., & Hulbert, L. G. (2007). Groupdrink: The effects of alcohol on intergroup competitiveness. *Psychology of Addictive Behaviors*, *21*(2), 272-276.

Horrace, W. C., & Oaxaca, R. L. (2006). Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters, 90*(3), 321-327.

Ishii, Y., Okubo, M., Nicholls, M. E., & Imai, H. (2011). Lateral biases and reading direction: A dissociation between aesthetic preference and line bisection. *Brain and Cognition*, *75*(3), 242-247.

Isoni, A., Poulsen, A., Sugden, R., & Tsutsui, K. (2013). Focal points in tacit bargaining problems: Experimental evidence. *European Economic Review*, *59*, 167-188.

Jackson, R. C., Warren, S., & Abernethy, B. (2006). Anticipation skill and susceptibility to deceptive movement. *Acta Psychologica*, *123*(3), 355-371.

Jiang, T. (2013). Cheating in mind games: The subtlety of rules matters. *Journal of Economic Behavior & Organization*, *93*, 328-336.

Jones, D. N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3). *Assessment*, *21*(1), 28-41.

Jordan, J. J., McAuliffe, K., & Warneken, F. (2014). Development of in-group favoritism in children’s third-party punishment of selfishness. *Proceedings of the National Academy of Sciences*, *111*(35), 12710-12715.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review, 76*(4), 728-741.

Kajackaite, A., & Gneezy, U. (2015). Lying costs and incentives. *Mimeo*.

Knoepfle, D. T., Wang, J. T.-Y., & Camerer, C. F. (2009). Studying learning in games using eye-tracking. *Journal of the European Economic Association*, *7*(2-3), 388-398.

Kocher, M. G., Schudy, S., & Spantig, L. (2017). I lie? We lie! Why? Experimental evidence on a dishonesty shift in groups. *Management Science*, (forthcoming).

Kovács-Bálint, Z., Bereczkei, T., & Hernádi, I. (2013). The telltale face: Possible mechanisms behind defector and cooperator recognition revealed by emotional facial expression metrics. *British Journal of Psychology*, *104*(4), 563-576.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience, 13*(10), 1292-1298.

Lahey, J. N., & Oxley, D. (2016). The power of eye tracking in economics experiments. *American Economic Review, 106*(5), 309-313.

Lane, S. D., Cherek, D. R., Pietras, C. J., & Tcheremissine, O. V. (2004). Alcohol effects on human risk taking. *Psychopharmacology, 172*(1), 68-77.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics, 1*(3), 593-622.

Levine, E. E., & Schweitzer, M. E. (2014). Are liars ethical? On the tension between benevolence and honesty. *Journal of Experimental Social Psychology, 53*, 107-117.

Levine, E. E., & Schweitzer, M. E. (2015). Prosocial lies: When deception breeds trust. *Organizational Behavior and Human Decision Processes, 126*, 88-106.

Levitt, S. D., List, J. A., & Reiley, D. H. (2010). What happens in the field stays in the field: Exploring whether professionals play minimax in laboratory experiments. *Econometrica, 78*(4), 1413-1434.

Litchfield, D., & Ball, L. J. (2011). Using another's gaze as an explicit aid to insight problem solving. *Quarterly Journal of Experimental Psychology, 64*(4), 649-656.

Litchfield, D., Ball, L. J., Donovan, T., Manning, D. J., & Crawford, T. (2010). Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *Journal of Experimental Psychology: Applied, 16*(3), 251-262.

Liu, C. (2008). *The Dark Forest* (J. Martinsen, Trans.). New York: Tor (translated version, 2015).

Lohse, T., & Qari, S. (2014). Gender differences in deception behaviour–The role of the counterpart. *Applied Economics Letters, 21*(10), 702-705.

MacDonald, T. K., Zanna, M. P., & Fong, G. T. (1995). Decision making in altered states: Effects of alcohol on attitudes toward drinking and driving. *Journal of Personality and Social Psychology, 68*(6), 973-985.

Mann, S. A., Vrij, A., Leal, S., Vernham, Z. L., & Geven, L. (2015). The deterrence of deception through imposing cognitive load. *Open Access Journal of Forensic Psychology, 7,* 90-105.

Marlowe, F. W., Berbesque, J. C., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Ensminger, J., Gurven, M., Gwako, E., & Henrich, J. (2008). More 'altruistic' punishment in larger societies. *Proceedings of the Royal Society of London B: Biological Sciences, 275*(1634), 587-592.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research, 45*(6), 633-644.

Milinski, M., & Rockenbach, B. (2012). On the interaction of the stick and the carrot in social dilemmas. *Journal of Theoretical Biology, 299,* 139-143.

Morewedge, C. K., Krishnamurti, T., & Ariely, D. (2014). Focused on fairness: Alcohol intoxication increases the costly rejection of inequitable rewards. *Journal of Experimental Social Psychology, 50*, 15-20.

Moskowitz, H., Burns, M. M., & Williams, A. F. (1985). Skills performance at low blood alcohol levels. *Journal of Studies on Alcohol, 46*(6), 482-485.

Muehlheusser, G., Roider, A., & Wallmeier, N. (2015). Gender differences in honesty: Groups versus individuals. *Economics Letters, 128*, 25-29.

Müller, R., Helmert, J. R., & Pannasch, S. (2014). Limitations of gaze transfer: Without visual context, eye movements do not help to coordinate joint action, whereas mouse movements do. *Acta Psychologica, 152*, 19-28.

Müller, R., Helmert, J. R., Pannasch, S., & Velichkovsky, B. M. (2013). Gaze transfer in remote cooperation: Is it always helpful to see what your partner is attending to? *Quarterly Journal of Experimental Psychology, 66*(7), 1302-1316.

Nagano, T., Kato, T., & Fukuda, T. (2006). Visual behaviors of soccer players while kicking with the inside of the foot. *Perceptual and Motor Skills, 102*(1), 147-156.

Nalanagula, D., Greenstein, J. S., & Gramopadhye, A. K. (2006). Evaluation of the effect of feedforward training displays of search strategy on visual search performance. *International Journal of Industrial Ergonomics, 36*(4), 289-300.

Neider, M. B., Chen, X., Dickinson, C. A., Brennan, S. E., & Zelinsky, G. J. (2010). Coordinating spatial referencing using shared gaze. *Psychonomic Bulletin & Review, 17*(5), 718-724.

Nikiforakis, N., & Mitchell, H. (2014). Mixing the carrots with the sticks: Third party punishment and reward. *Experimental Economics, 17*(1), 1-23.

Norton, E. C., Wang, H., & Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. *Stata Journal, 4*(2), 154-167.

Penczynski, S. P. (2016). Strategic thinking: The influence of the game. *Journal of Economic Behavior & Organization, 128*, 72-84.

Peterson, J. B., Rothfleisch, J., Zelazo, P. D., & Pihl, R. (1990). Acute alcohol intoxication and cognitive functioning. *Journal of Studies on Alcohol, 51*(2), 114-122.

Polonio, L., Di Guida, S., & Coricelli, G. (2015). Strategic sophistication and attention in games: An eye-tracking study. *Games and Economic Behavior, 94*, 80-96.

Potters, J., & Stoop, J. (2016). Do cheaters in the lab also cheat in the field? *European Economic Review, 87*, 26-33.

Proestakis, A., Espín, A. M., Exadaktylos, F., Cortés Aguilar, A., Oyediran, O. A., & Palacio, L. A. (2013). The separate effects of self-estimated and actual alcohol intoxication on risk taking: A field experiment. *Journal of Neuroscience, Psychology, and Economics, 6*(2), 115-135.

Pruckner, G. J., & Sausgruber, R. (2013). Honesty on the streets: A field study on newspaper purchasing. *Journal of the European Economic Association, 11*(3), 661-679.

Raab, M., & Gigerenzer, G. (2015). The power of simplicity: A fast-and-frugal heuristics approach to performance science. *Frontiers in Psychology, 6*(1672), 1-6.
Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review, 83*(5), 1281-1302.
Rapoport, A., & Budescu, D. V. (1992). Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General, 121*(3), 352-363.
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59-108.
Rivers, D., & Vuong, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics, 39*(3), 347-366.
Rowe, R., Horswill, M. S., Kronvall-Parkinson, M., Poulter, D. R., & McKenna, F. P. (2009). The effect of disguise on novice and expert tennis players' anticipation ability. *Journal of Applied Sport Psychology, 21*(2), 178-185.
Rubinstein, A., Tversky, A., & Heller, D. (1997). Naive strategies in competitive games. In W. Albers, W. Güth, P. Hammerstein, B. Moldovanu, & E. van Damme (Eds.), *Understanding Strategic Interaction: Essays in Honor of Reinhard Selten* (pp. 394-402). Berlin, Heidelberg: Springer.
Ruffle, B., & Tobol, Y. (2017). Clever enough to tell the truth. *Experimental Economics, 20*(1), 130-155.
Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, 71-78.
Sayette, M. A. (2017). The effects of alcohol on emotion in social drinkers. *Behaviour Research and Therapy, 88*, 76-89.
Schelling, T. C. (1960). *The Strategy of Conflict*. Cambridge, MA; London, England: Harvard University Press (reprinted and edited version, 1980).
Schilbach, F. (2015). Alcohol and self-control: A field experiment in India. *Mimeo*.
Schreiber Compo, N., Evans, J. R., Carol, R. N., Kemp, D., Villalba, D., Ham, L. S., & Rose, S. (2011). Alcohol intoxication and memory for events: A snapshot of alcohol myopia in a real-world drinking scenario. *Memory, 19*(2), 202-210.
Schwedes, C., & Wentura, D. (2012). The revealing glance: Eye gaze behavior to concealed information. *Memory & Cognition, 40*(4), 642-651.
Sebanz, N., & Shiffrar, M. (2009). Detecting deception in a bluffing body: The role of expertise. *Psychonomic Bulletin & Review, 16*(1), 170-175.
Sefton, M., Shupp, R., & Walker, J. M. (2007). The effect of rewards and sanctions in provision of public goods. *Economic Inquiry, 45*(4), 671-690.

Shalvi, S. (2012). Dishonestly increasing the likelihood of winning. *Judgment and Decision Making, 7*(3), 292-303.

Shalvi, S., & De Dreu, C. K. (2014). Oxytocin promotes group-serving dishonesty. *Proceedings of the National Academy of Sciences, 111*(15), 5503-5507.

Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological Science, 23*(10), 1264-1270.

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience, 6*(12), 1317-1322.

Sjöberg, L. (1969). Alcohol and gambling. *Psychopharmacology, 14*(4), 284-298.

Stein, R., & Brennan, S. E. (2004). Another person's eye gaze as a cue in solving programming problems. *Proceedings of the 6th International Conference on Multimodal Interface*, 1-7.

Stepner, M. (2014). *Binned Scatterplots: Introducing binscatter and exploring its applications.* Paper presented at the 2014 Stata Conference.

Stewart, N., Gächter, S., Noguchi, T., & Mullett, T. L. (2016). Eye movements in strategic choice. *Journal of Behavioral Decision Making, 29*(2-3), 137-156.

Sutter, M. (2009). Deception through telling the truth?! Experimental evidence from individuals and teams. *The Economic Journal, 119*(534), 47-60.

Sutter, M., Haigner, S., & Kocher, M. G. (2010). Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *Review of Economic Studies, 77*(4), 1540-1566.

Tay, C. S., Chow, J. Y., Koh, M., & Button, C. (2010). The effectiveness of keeper-independent penalty kicks using fake visual cues from penalty takers. *International Journal of Sport Psychology, 43*(5), 403-419.

Turillo, C. J., Folger, R., Lavelle, J. J., Umphress, E. E., & Gee, J. O. (2002). Is virtue its own reward? Self-sacrificial decisions for the sake of fairness. *Organizational Behavior and Human Decision Processes, 89*(1), 839-865.

van 't Veer, A., Stel, M., & van Beest, I. (2014). Limited capacity to lie: Cognitive load interferes with being dishonest. *Judgment and Decision Making, 9*(3), 199-206.

Van Leeuwen, B., Noussair, C. N., Offerman, T., Suetens, S., van Veelen, M., & van de Ven, J. (2017). Predictably angry–Facial cues provide a credible signal of destructive behavior. *Management Science*, (online first).

Velichkovsky, B. M. (1995). Communicating attention: Gaze position transfer in cooperative problem solving. *Pragmatics & Cognition, 3*(2), 199-223.

Vrij, A. (2004). Why professionals fail to catch liars and how they can improve. *Legal and Criminological Psychology, 9*(2), 159-181.

Walker, J. M., & Halloran, M. A. (2004). Rewards and sanctions and the provision of public goods in one-shot settings. *Experimental Economics, 7*(3), 235-247.

Walker, M., & Wooders, J. (2001). Minimax play at Wimbledon. *American Economic Review, 91*(5), 1521-1538.

Wang, J. T.-Y., Spezio, M., & Camerer, C. F. (2010). Pinocchio's pupil: Using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games. *American Economic Review, 100*(3), 984-1007.

Weisel, O., & Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences, 112*(34), 10651-10656.

Wermeskerken, M., Litchfield, D., & Gog, T. (2017). What Am I Looking at? Interpreting Dynamic and Static Gaze Displays. *Cognitive Science,* (online first).

Wiltermuth, S. S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes, 115*(2), 157-168.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (2. ed.). Cambridge, Massachusetts: MIT Press.