I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11031

Partners in Crime: Diffusion of Responsibility in Antisocial Behaviors

Sascha Behnk
Li Hao
Ernesto Reuben

SEPTEMBER 2017

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11031

Partners in Crime: Diffusion of Responsibility in Antisocial Behaviors

Sascha Behnk
University of Zurich

Li Hao
University of Arkansas

Ernesto Reuben
New York University Abu Dhabi and IZA

SEPTEMBER 2017

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Partners in Crime: Diffusion of Responsibility in Antisocial Behaviors

Using a series of sender-receiver games, we find that two senders acting together are willing to behave more antisocially towards the receiver than single senders. This result is robust in two contexts: when antisocial messages are dishonest and when they are honest but unfavorable. Our results suggest that diffusion of responsibility is the primary reason for the increased antisocial behavior as our experimental design eliminates competing explanations. With a partner in crime, senders think that behaving antisocially is more acceptable and experience less guilt. Importantly, we identify a crucial condition for the increased antisocial behavior by groups: the partner in crime must actively participate in the decision-making. Our results have important implications for institutional design and promoting prosocial behaviors.

JEL Classification: D70, D91, C92, D63

Keywords: diffusion of responsibility, antisocial behavior, moral norms, guilt aversion

Corresponding author:
Ernesto Reuben
New York University Abu Dhabi
PO Box 129188
Abu Dhabi
United Arab Emirates

E-mail: ereuben@nyu.edu

1. Introduction

Does the willingness of individuals to behave prosocially depend on whether decisions are made individually or collectively? An increasing amount of evidence suggests that people are more likely to behave selfishly or antisocially when acting together than when acting alone.[1] A compelling explanation for why joint decision-making decreases prosocial behavior is diffusion of responsibility (Dana, Weber, and Kuang 2007; Bartling, Fischbacher, and Schudy 2015), which we define as a reduction in the intrinsic disutility individuals incur from acting antisocially because more people are involved in the decision-making. In this paper, we experimentally study such diffusion of responsibility in two types of antisocial acts. More specifically, we answer the following questions. Does knowing that other people are involved in the decision-making generate diffusion of responsibility even when individuals do not communicate nor interact in any form? Is diffusion of responsibility the result of individuals perceiving their antisocial behavior as more acceptable and therefore suffering less when acting antisocially? Is the mere presence of another person enough to trigger these effects or is that person required to be actively involved in the decision-making?

We answer these questions by studying sender-receiver games in a laboratory experiment. In all games, an uninformed receiver chooses one out of ten options to determine the players' earnings. The receiver's only information is a message that is either prosocial (identifying the option that gives everyone an equal payoff) or antisocial (identifying the option that benefits sender(s) at the expense of receivers).[2] We study two types of antisocial messages: an antisocial message that reveals the unequal outcome and truthfully indicates it as such, and a deceptive antisocial message that points to the unequal outcome but claims it is revealing the equal outcome. We compare a game where the message is chosen by a single sender to a game where the message is chosen by two senders who cannot communicate with each other. Importantly, we use a price list to elicit the minimum monetary

[1] For example, increased selfish behavior with "partners in crime" has been observed in several contexts, including altruistic giving (Dana, Weber, and Kuang 2007), reciprocity (Cox 2002; Kocher and Sutter 2007), lying (Cohen et al. 2009; Sutter 2009; Weisel and Shalvi 2015; Barr and Michailidou 2017; Kocher, Schudy, and Spantig 2017; Nielsen et al. 2017), and markets of goods with negative externalities (Falk and Szech 2013; Bartling, Weber, and Yao 2015).

[2] The remaining eight options are Pareto-dominated and pay all players a much smaller amount.

compensation required by each sender to send the antisocial message, which we call the "antisocial premium". In short, our experimental design offers a precise measure of each sender's willingness to act antisocially, which allows us to test the extent to which diffusion of responsibility occurs in two different types of antisocial behaviors.

There are numerous reasons why collective decisions can result in more antisocial outcomes than individual decisions. For example, Falk and Szech (2013) and Kocher, Schudy, and Spantig (2017) propose the following potential effects: (i) diffused responsibility and diminished sense of guilt due to the presence of other decision-makers; (ii) learning through the interaction with others that some are willing to ignore ethical norms decreases one's willingness to abide by those norms; (iii) arguments made during group discussions justifying antisocial behavior might be more effective than those promoting prosocial behavior; (iv) direct effects of the specific rules used to aggregate individual preferences (e.g., majority vs. unanimity), and (v) the introduction of materialistic framing embedded in some institutions used to make the collective decisions (e.g., markets) might divert attention away from the acceptability of the antisocial behavior.[3] In this paper, we concentrate solely on diffusion of responsibility and exclude the other effects through our experimental design. Specifically, by eliminating all interaction between senders, we exclude any learning or argumentation effects, and by using the same price list to elicit all senders' antisocial premiums, we obtain a comparable measure of the senders' preferences that is independent of the institution used to determine the final outcome.

We find that senders' antisocial premiums are significantly lower when there are two senders compared to when there is only one sender. Moreover, the difference in antisocial premiums is present and of similar magnitude irrespective of whether the antisocial message is deceptive or not. This finding demonstrates that the simply presence of another decision-maker is already enough to lower the willingness to act prosocially.

We further study whether the difference in antisocial premiums is in fact related to changes in the disutility senders experience when sending the antisocial message to the receiver. To do so, we

[3] For example, if the collective decision involves bargaining, then the individuals' attention might shift away from moral considerations as individuals rather focus on the bargaining process than on the moral implications of the potential outcomes (see Falk and Szech 2013).

obtain two measures of how unpleasant sending the antisocial message is. First, we ask subjects to rate the acceptability of sending the antisocial message, which combined with the subjects' beliefs about their actions gives as a general measure of the expected disutility of behaving antisocially.[4] Second, we use self-reports to directly measure how much guilt senders' experience when they send the antisocial message.[5] This measure allows us to confirm whether the senders' hedonic experience following their decisions is consistent with their behavior and normative evaluations.

When we analyze the subjects' normative ratings, we find that subjects indeed think that sending the antisocial message is significantly more acceptable when there are two senders instead of one. Importantly, this view is held by both senders and receivers, suggesting that the difference in ratings is not merely the result of senders' internal justification of their own choices. Analyzing the relationship between antisocial premiums and acceptability ratings reveals that diffusion of responsibility is best explained by a change in the relationship between second-order beliefs and antisocial premiums among senders who are already inclined to think that the antisocial message is not too unacceptable. The senders' ex post emotional experience is also revealing. We find that senders feel less guilt when they send the antisocial message in the two-sender game than in the one-sender game. In other words, the effect of a second sender is not only evident in the senders' behavior but also in their emotional state.

A natural question that arises in this context is whether the difference in the antisocial premiums is due to the fact that the outcome affects a second individual or whether the active participation of that individual in the decision-making is necessary. We address this question by introducing a third sender-receiver game in which there is a passive sender who receives the same payoffs as the second sender in the two-sender games but does not have say in which message is sent. The answer to this question is not only relevant from a policy perspective, it also allows us to narrow down the type of models that can be used to explain the difference in antisocial premiums.

---

[4] We opted for a general method to measure the subjects' disutility instead of assuming specific functional forms because there are many models of social preferences and there is no consensus in the literature as to which is the best model to use (see Fehr and Schmidt 2006).

[5] Social emotions like guilt has been shown to be a crucial determinants of whether individuals behave prosocially or antisocially (van Winden 2007; Hopfensitz and Reuben 2009; Reuben and van Winden 2010).

We find that antisocial premiums with a passive second sender are higher than when the second sender is active and are very similar to the antisocial premiums in one-sender games. This result shows that the active involvement of the second sender in the decision-making process is crucial for diffusion of responsibility to occur.

Our contribution is threefold. First, we demonstrate senders act more antisocially with a partner in crime than when they act alone. We argue that this is solely due to diffusion of responsibility caused by the presence of a second decision-maker, as our experimental design eliminates explanations the require interaction within the group. Second, with a partner in crime, senders think that antisocial messages are normatively more acceptable, and they experience less guilt when the send such messages, which are in line with diffusion of responsibility. Finally, we identify a crucial condition for the effect of a partner in crime: the partner must actively participate in the decision-making. In other words, responsibility is not diffused with a passive partner who does not partake in the decision-making process.

2. Related literature

This paper builds on previous studies that investigate how interacting with others affects one's proclivity to act antisocially. Diffusion of responsibility in this regard has been mainly presumed in two circumstances: when individuals interact by bargaining over an antisocial action in a market and when decisions are jointly made in groups.[6]

2.1. Diffusion of responsibility via market interactions

Falk and Szech (2013) show that subjects are more inclined to accept the death of a mouse in return for money when they bargain over the respective amount in a bilateral or multilateral double-auction market than when they make their decisions individually. Kirchler et al. (2016) confirm this trend by testing how different interventions concerning morality affect donation decisions via individual

[6] Increased antisocial behavior due to the involvement of others is also observed in settings where responsibility is transferred through delegation or intermediation. For example, delegation leads to lower payoffs for receivers in dictator games (Hamman, Loewenstein, and Weber 2010; Bartling and Fischbacher 2012) and intermediation leads to less third-party punishment of unfair allocations resulting in more selfish decisions (Coffman 2011; Oexl and Grossman 2013; Garofalo and Rott 2017).

choice lists and double-auction markets. The assumption that market environments lead to an erosion of morality is further supported by Deckers et al. (2016), who find that personal characteristics that predict antisocial behavior when decisions are made individually lose their predictive power in market settings. Bartling, Weber, and Yao (2015) examine the willingness to maximize own payoffs at the cost of a third player in markets with subjects from countries with low (China) and high (Switzerland) degrees of market-orientation. They find that subjects are less socially responsible in market settings compared to individual decisions and that this pattern is more pronounced among subjects with a lower degree of market-orientation.

Unlike these studies, we employ a setting where information about the normative views of others is not revealed through trade and without a market framing that can distract subjects from thinking about the acceptability of the antisocial outcome. In other words, we concentrate solely on the effect of the inclusion of another decision-maker on the willingness of individuals to behave antisocially.

2.2. Diffusion of responsibility via group decisions

The second line of literature investigates whether, compared to individuals, behavior is more antisocial when decisions are made in groups. Evidence from multiple economic games provides only mixed support to this assertion.

Various studies have found increased antisocial behavior in groups. In the context of dictator games, Dana, Weber, and Kuang (2007) find that two individuals making a joint decision give less than single dictators. Moreover, there is some evidence that groups act more selfishly as proposers in ultimatum games (Bornstein and Yaniv 1998), as trustees in trust games (Cox 2002), and as workers in gift-exchange games (Kocher and Sutter 2007). In the context of deception and lying, Keck (2014) provides evidence that receivers who make joint decisions in an ultimatum game are more likely to deceive the proposer than single receivers. Kocher, Schudy, and Spantig (2017) find that groups are more willing to lie than individuals in a die-rolling game (Fischbacher and Föllmi-Heusi 2013), even when group members are paid only according to their own decisions.[7] Other variations

[7] Several studies report increased dishonesty when lies benefit others through aligned payoffs—both when decisions are made individually (Wiltermuth 2011; Gino and Pierce 2010; Conrads et al. 2013) and jointly (Gino, Ayal, and Ariely 2013;

of the anonymous die-rolling game with joint decisions provide further evidence of comparatively more dishonest acts by groups (Gino, Ayal, and Ariely 2013; Chytilová and Korbel 2014; Muehlheusser, Roider, and Wallmeier 2015; Weisel and Shalvi 2015). Nielsen et al. (2017) find that teams are less likely to keep promises made as trustees in a trust game. Finally, in sender-receiver games (Gneezy 2005), Sutter (2009) finds that groups are more likely to use strategic deception by telling the truth when they expect receivers will not follow their message, while Cohen et al. (2009) find that senders making joint decisions deceive receivers more often than individual senders when they are informed of the receiver's decision ahead of time.

However, there are also other studies that find less antisocial behavior by groups than by individuals. For example, Cason and Mui (1997) find more selfish behavior by dictators when they are individuals as opposed to groups (but see the critique by Luhan, Kocher, and Sutter 2009). In sender-receiver games, in addition to the results reported above, Sutter (2009) also finds that groups of senders lie less than single senders when they expect receivers will follow their message, and Cohen et al. (2009) find that groups deceive receivers less than individuals when the receivers' behavior is unknown at the time the decision is made.

When comparing individuals to groups there are reasons other than diffusion of responsibility that could potentially explain why groups can act more antisocially. By using the strategy method to elicit the senders' willingness to act antisocially and eliminating all communication between decision-makers, we can observe the senders' preferences independently of the preference-aggregation rules used in the group decision-making process and we exclude any peer or argumentation effects.

There are two other studies that investigate diffusion of responsibility in antisocial behaviors by comparing a single decision-maker to multiple decision-makers who cannot communicate with each other.[8] In one of their treatments, Dana, Weber, and Kuang (2007) compare dictator giving between

---

Weisel and Shalvi 2015). Danilov et al. (2013) find that aligned payoffs result in more dishonesty only when social ties are strong among players.

[8] Barr and Michailidou (2017) use a die-rolling game to study the willingness of individuals to lie in order to benefit another player, knowing that the other player is making the same lying decision to potentially benefit them. They find more lying when individuals make this lying decision compared to a control where the beneficiary of the lie is a passive player.

single dictators and two dictators who simultaneously submit a giving decision knowing that the unfair outcome will be implemented only if they both choose it. They find more selfish actions with two dictators. Relatedly, Falk and Szech (2016) find that the fraction of unethical decisions (whether to let a mouse be killed in return for a certain payoff) is higher when subjects are grouped and thus cannot find out whether their decision was pivotal for the killing. Our study is different from these studies in three respects. First, the use of the strategy method allows us to get a more precise measure of the extent to which diffusion of responsibility occurs. Second, and more importantly, we also obtain measures of the subjects' normative views, beliefs, and emotions in order to test directly whether and how diffusion of responsibility occurs. Third, our experimental design allows us to investigate an important question in this regard: does diffusion of responsibility require the active involvement of others in the decision-making process?

3. Sender-receiver games

We employ a two-by-two experimental design using variations of sender-receiver games. The games vary in the number of senders, either one or two, and the type of antisocial message senders can deliver, either truthful or deceptive.

In our *1-Sender* games, subjects are randomly divided into groups of two and assigned either the role of a sender or a receiver, neutrally named Player 1 and Player 2. The receiver's task is to choose one out of ten options in order to determine both players' earnings.

There is one prosocial option that pays €10 to each player, one antisocial option that pays the sender €17 minus an amount $x \in$ [€0, €6.5] and €3 to the receiver, and eight Pareto-dominated options that pay €4 to the sender and €0 to the receiver. At the beginning of the game, the computer randomly labels each of the ten options with a unique letter ranging from A to J. Although both players know the payoff consequences of a particular option being chosen, only the sender knows how each of the ten options is labeled. Table 1(a) is an example of a letter assignment and how this information is presented to the sender.

Although related, this study is different from ours and others in this literature in that there is no joint antisocial decision to be made.

Table 1. Example of payoff tables in the sender-receiver games (amounts in euros)

(a) *1-Sender* game

Option	A	B	C	D	E	F	G	H	I	J
Sender	4	4	10	4	17 – x	4	4	4	4	4
Receiver	0	0	10	0	3	0	0	0	0	0

(b) *2-Sender* game

Option	A	B	C	D	E	F	G	H	I	J
Sender A	4	4	10	4	17 – x	4	4	4	4	4
Sender B	4	4	10	4	10 + x	4	4	4	4	4
Receiver	0	0	10	0	3	0	0	0	0	0

In the *2-Sender* game, subjects are randomly matched in groups of three and are assigned either the role of sender A, sender B, or receiver, neutrally framed as Players 1, 2 and 3. The payoff structure for sender A and the receiver are identical to those of the sender and receiver in the *1-Sender* game. The new player, sender B, receives identical payoffs as sender A in all nine options except in the antisocial option where sender B receives €10 plus the amount x, as seen in Table 1(b).

The only information available to the receiver regarding the label assignment of the ten options is due to a message. In the *1-Sender* game, the sender decides which message is sent to the receiver, whereas in the *2-Sender* game, sender A and sender B jointly make this decision. There are two available messages. The first message, Message I, accurately reveals the label of the prosocial option and reads "Option <letter paying the receiver €10> will earn you 10 euros". The second message, Message II, is one of two types, depending on the "context". In the *Bitter pill* context, Message II accurately reveals the label of the antisocial option and reads "Option <letter paying the receiver €3> will earn you 3 euros". In the *Deception* context, Message II is deceptive in that it reveals the label of the antisocial option but claims it is the label of the prosocial option: "Option <letter paying the receiver €3> will earn you 10 euros". Like senders, receivers are aware that there are two available messages and that, in *Deception*, a message can be deceptive. Hence, it is common knowledge that a message always reveals the label of either the prosocial or the antisocial option and never the label of one of the Pareto-dominated options.

Our aim with this design is to let receivers make an informed decision in order for them to have well-defined beliefs about the senders' behavior (in contrast to papers based on the design of Gneezy

2005; see Sutter 2009) whilst maintaining the senders' incentive to reveal their preferences. In other words, we selected the payoffs and number of Pareto-dominated options to ensure that enough receivers follow the message for senders to have an incentive to choose the message corresponding to their preferred outcome.[9]

3.1. The antisocial premium

We use the strategy method to measure the senders' willingness to send an antisocial message. Specifically, senders in the *1-Sender* game as well as senders A in the *2-Sender* game choose between Message I and Message II in each of the 14 rows in Table 2(a). Senders B in the *2-Sender* game choose between Message I and Message II in each of the 14 rows in Table 2(b). Senders A and B make their decisions simultaneously. Thereafter, the computer randomly selects one row to determine which message is sent.[10] In the *2-Sender* game, Message II is sent only if both senders choose Message II, otherwise Message I is sent. In other words, the antisocial message is sent only with the consent of both senders. Importantly, this procedure ensures that senders in the *2-Sender* game always have a (weakly) positive incentive to choose according to their preferences.

While Message I always pays €10, the payoff of Message II depends on the amount x. By systematically varying x, we measure the minimum monetary compensation senders must receive in order to send the antisocial Message II instead of the prosocial Message I. Accordingly, we call this minimum compensation the senders' *antisocial premium*. More specifically, senders who choose Message II over Message I for a given x are classified as having an antisocial premium in the interval [€x – 0.5, €x].[11]

---

[9] Specifically, if we define p as the probability that the receiver follows the message, then a sender who prefers option m to option n will choose the message corresponding to m as long as $pU(m) + (1 - p)(1/9)U(n) + (1 - p)(8/9)U(d) > pU(n) + (1 - p)(1/9)U(m) + (1 - p)(8/9)U(d)$, where $U(.)$ is the sender's utility if the option implemented is m, n, or d (a dominated option). It is easy to see that this inequality holds as long as $p > 1/9$. Given the receivers' incentives, even if they possess social preferences, it is highly unlikely that only 11% of them follow the message.

[10] When receivers see the message, they are not informed which row was selected by the computer.

[11] At the extremes, senders who always choose Message I are classified as having an antisocial premium in the interval [€7.50, ∞) if they played in the *1-Sender* game or as sender A in the *2-Sender* game, or in the interval [€7.00, ∞) if they played as sender B in the *2-Sender* game. The analogous intervals for senders who always choose Message II are (-∞, €0.50] and (-∞, €0.00]. Senders who switched more than once or switched in the wrong direction were not classified.

Table 2. Senders' choice lists (amounts in euros)

(a) Sender (*1-Sender*) and sender A (*2-Sender*)

Row	Payoff of Message I	x	Payoff of Message II
1	10	0.0	17.0
2	10	0.5	16.5
3	10	1.0	16.0
4	10	1.5	15.5
5	10	2.0	15.0
6	10	2.5	14.5
7	10	3.0	14.0
8	10	3.5	13.5
9	10	4.0	13.0
10	10	4.5	12.5
11	10	5.0	12.0
12	10	5.5	11.5
13	10	6.0	11.0
14	10	6.5	10.5

(b) Sender B (*2-Sender*)

Row	Payoff of Message I	x	Payoff of Message II
1	10	0.0	10.0
2	10	0.5	10.5
3	10	1.0	11.0
4	10	1.5	11.5
5	10	2.0	12.0
6	10	2.5	12.5
7	10	3.0	13.0
8	10	3.5	13.5
9	10	4.0	14.0
10	10	4.5	14.5
11	10	5.0	15.0
12	10	5.5	15.5
13	10	6.0	16.0
14	10	6.5	16.5

3.2. Hypotheses

In this subsection, we present testable hypotheses to guide our data analysis. Prior empirical literature suggests that antisocial premiums will be higher in the *1-Sender* compared to the *2-Sender* games. From a theoretical perspective, however, it is unclear why this might be the case.

The standard model of own-payoff maximization predicts that senders always choose the antisocial Message II and receivers pick the option indicated in the message in both the *1-Sender* and *2-Sender* games. However, given the abundant amount of evidence from laboratory experiments and the field, it is reasonable to expect that a majority of subjects will not behave this way. In this respect, it is more interesting to consider models that assume people are motivated by more than just monetary payoffs and predict there are individuals with positive antisocial premiums. To simplify this discussion, we assume that enough receivers follow the message such that senders have a dominant strategy to choose the message corresponding to their preferred outcome. This assumption is not restrictive, as we discuss in footnote 9, and is borne out by the data since only 6 out of all 79 receivers deviated from the recommended option.

In many models of social preferences, senders experience disutility if the send the antisocial Message II. For example, outcome-based models where senders experience disutility because they care about the earnings of the receiver (Andreoni 1990), the difference in earnings between themselves and the receiver (Fehr and Schmidt 1999; Bolton and Ockenfels 2000), their reputation or self-image as generous individuals (Bénabou and Tirole 2006), the welfare of the least well off (Charness and Rabin 2002), or their honesty (Ellingsen and Johannesson 2004). Also, intention based models where senders experience disutility when they act unkindly (Rabin 1993) or if they disappoint the expectations of the receiver (Battigalli and Dufwenberg 2007). All these models allow for senders with positive antisocial premiums. However, do they predict that the disutility experienced by sending Message II is reduced with the inclusion of a second sender?

Broadly speaking, models of social preferences can be organized into three categories to explain why antisocial premiums may decrease from the *1-Sender* to the *2-Sender* games. The first category includes outcome-based models where other-regarding concerns are weighted in some way by the number of players. In these models, the inclusion of a second sender implies that the weight associated with the receiver is lowered. As an example, consider the model of Bolton and Ockenfels (2000) where senders experience increasing disutility the more their income deviates from the average income. In the *1-Sender* game, sending Message II implies a large income difference between the sender and the average income. In the *2-Sender* game, sending Message II implies a smaller income difference between the sender and the average income simply because the average income now includes the income of the second sender who also benefits from sending Message II. Similar effects are seen in the models of Fehr and Schmidt (1999) and Charness and Rabin (2002).

The second category includes intention-based models of social preferences, which are characterized by the players' utility being directly affected by the beliefs of other players. These models are generally complex and result in multiple equilibria that depend on the players' beliefs. However, it is straightforward to intuit that models like guilt aversion (Battigalli and Dufwenberg 2007), where senders feel guilt from sending Message II if the receiver expects them to send Message

I, imply lower antisocial premiums in the *2-Sender* game than the *1-Sender* game if the inclusion of the second sender increases the receiver's belief that they will receive Message II.

Lastly, the third category comprises of models that have been specifically created to model diffusion of responsibility by assuming that an individual's utility of an antisocial outcome depends on how likely it is that that individual's choice was pivotal in determining the outcome (Engl 2017; Rothenhäusler, Schweizer, and Szech 2017). These models generally predict that increases in the number of decision makers result in more antisocial behavior.[12] Based on these arguments, we formulate our first hypothesis:

Hypothesis 1: *On average, antisocial premiums are lower in 2-Sender games than 1-Sender games.*

In addition to studying diffusion of responsibility through decisions, we also investigate subjects' beliefs and emotions. Our goal is to test whether differences in antisocial premiums, both across games and across subjects, are indeed the result of differences in the subjects' correctly anticipated disutility of sending Message II.

We elicit the senders' normative views regarding the prosocial and antisocial messages and use them as a general measure of their expected disutility of behaving antisocially due to other-regarding concerns. Specifically, after they made their decisions and delivered the message, but before they learned the final outcome, we asked senders to indicate for each message "How acceptable do you consider it is to deliver Message I [or Message II] to Player 3 [the receiver]?" Answers were recorded with a 5-point Likert scale ranging from "very unacceptable" (1) to "very acceptable" (5). We opted for a simple normative-views elicitation because there is no consensus yet on the most relevant model of social preferences (see Fehr and Schmidt 2006), which means that concentrating on a specific functional form entails a large misspecification risk. With our approach, we simply need to assume that senders will rate the acceptability of each message according their specific other-regarding motivations. For example, we are assuming that senders who care a lot about efficiency and senders who care a lot about their self-image will both rate sending Message I as being much

[12] Given that senders can unilaterally choose Message I in both the *1-Sender* and the *2-Sender* games, these models do not clearly imply diffusion of responsibility in our experiment. However, we include these models in the discussion because they attempt to define more precisely the meaning of responsibility, which is how the empirical literature generally motivates the difference between individual and collective decision-making.

more acceptable than sending Message II. This more general approach has been found to work well in various settings both in the laboratory (Bicchieri and Xiao 2009; Krupka and Weber 2013; Reuben and Riedl 2013; Erkut, Nosenzo, and Sefton 2015) and the field (Bicchieri 2017).

In addition to the senders, we also asked receivers to rate the acceptability of sending each message. Receivers were asked after the game was played and they had made their choice but before they learned their final earnings. Eliciting the receivers' normative views is interesting for two reasons. First, they give us a benchmark to evaluate wither the normative views of senders are self-serving. Second, it allows us to evaluate whether any *differences* in the senders' normative views between the *1-Sender* and *2-Sender* games are self-serving. In other words, they allow us to test whether diffusion of responsibility is specific to the senders' normative perceptions or whether it is a more general perceptual change. Finally, we also elicited the senders' belief of the receivers' normative views. To do so, after indicating their own normative views, we showed senders the question used to measure the normative views of the receiver in their group, and we asked them to indicate "What do you think was Player 3's [the receiver's] answer to this question?" We incentivized their answer by paying them €0.25 if they guessed correctly. The senders' belief of the receivers' normative views is interesting in that it tells us whether senders think that a second sender affects only their own other-regarding concerns or whether it also affects the other-regarding concerns of others.

In addition to their normative views, which are an anticipatory measure of the senders' utility of sending each message, we further study the disutility of sending each message by measuring the senders' experienced guilt after the message is actually sent. This measure allows us to confirm whether the senders' hedonic experience when their decision is implemented is consistent with their previous normative evaluations and whether it varies between the *1-Sender* and *2-Sender* games. To measure the senders' guilt, we asked them to self-report the intensity at which they experienced a series of emotions (i.e., guilt, shame, anger, happiness, and gratitude), each on a 7-point Likert scale that ranged from "not at all" (1) to "very intensively" (7). We asked subjects to report their emotions in the moment they saw the option that was implemented by the receiver and the earnings of all the players they were matched with. We use self-reported measures of emotions because we are particularly interested in measuring guilt. This emotion is known to be important in deterring individuals from acting antisocially and it is often used to motivate models of social preferences. To the best of our knowledge, there is no clear physiological measure of guilt (Adolphs 2002). This is not to say that self-reports do not have limitations. In particular, one might worry that subjects do not report their true emotions and instead report a fictitious emotional reaction. Reassuringly, considerable research has demonstrated that self-reported emotional experiences are highly correlated with various physiological measures like heart rates, facial movements, and brain activation (Bradley and Lang 2000; Ben-Shakhar et al. 2007).

If diffusion of responsibility occurs and the senders' disutility from sending Message II is lower in the *2-Sender* games than in the *1-Sender* games then the following hypotheses follow:

Hypothesis 2a: *On average, senders rate sending the antisocial message to the receiver as more acceptable in the 2-Sender game than in the 1-Sender game.*

Hypothesis 2b: *On average, if the receiver implements the antisocial message, senders experience lower intensities of guilt in the 2-Sender game than in the 1-Sender game.*

As discussed above, beliefs are direct determinants of one's willingness to act antisocially in intention-based models of social preferences. Although beliefs are incorporated into the utility function of individuals in different ways, an approach that seems particularly relevant to our game is that of models of guilt aversion (Battigalli and Dufwenberg 2007; Battigalli, Charness, and Dufwenberg 2013). In these models, the beliefs that trigger guilt are the senders' second-order beliefs. Specifically, senders experience more guilt for sending Message II if they believe they are disappointing the receiver, i.e., if they believe that the receiver's expected probability of receiving Message II is low. Hence, these models predict a negative relationship between the senders' second-order beliefs and antisocial premiums.[13]

To have a direct test of these predictions, we elicit the senders' beliefs about the receivers' expected probability of receiving Message II. In order to do so, we first elicit the receivers' expected

---

[13] In fact, if we take the normative views as a measure of the guilt sensitivity parameter (i.e., how much guilt one feels for a given second-order belief), then models of guilt aversion predict that antisocial premiums are the result of the interaction between normative views and second-order beliefs. For example, the senders with the highest antisocial premiums would be senders who think sending Message II is very unacceptable and believe receivers expect to receive Message I.

probability of receiving Message II by asking them "How many of the Players 1 [or pairs of Players 1 and 2, which are senders] in the room do you think will deliver Message II?" Thereafter, we showed this question to the senders and asked them to indicate "What do you think was Player 3's [the receiver's] answer to this question?" These elicitations were incentivized by paying senders €0.25 and receivers €0.75 per correct answer.[14] One way for models of guilt aversion to explain diffusion of responsibility is if second-order beliefs to vary between the *2-Sender* than in the *1-Sender* game according to our last hypothesis:

Hypothesis 3: *On average, the senders' belief of the receivers' expected probability of receiving the antisocial message is higher in the 2-Sender game compared to the 1-Sender game.*

3.3. Procedures

We ran the experiment between February and March 2015 at the Laboratory of Experimental Economics (LEE) at University Jaume I in Castellon, Spain. A total of 197 undergraduate students including 112 men and 85 women from different faculties were recruited using ORSEE (Greiner 2015). We conducted eight sessions, each lasting around 1.5 hours.

Upon arrival, the subjects were randomly assigned to computer terminals. Thereafter, the instructions for the experiment were read aloud by the experimenter and subjects were asked to answer a series of control questions (instructions are available in the Appendix). Subjects had the possibility to ask questions during the whole process. The experiment was then conducted using z-Tree (Fischbacher 2007).

Once senders made a decision for each of the 14 values of *x* (see Table 2), the computer randomly selected one of these values and displayed the text of the chosen message on the senders' screen. All senders in the *1-Sender* game and Sender B in the *2-Sender* game were asked to write down the message on an empty sheet of paper located on their desk and then wait for an experimenter to arrive to their desk. There were three experimenters in each session. The experimenter checked whether the written message coincides with the text on the screen and then guided the sender to their

[14] Unbeknownst to the subjects, we paid receivers a higher amount for an accurate answer in order to partly compensate them for their potentially lower earnings. We also elicited the sender's beliefs about the receivers' behavior by asking them "How many of the Players 3 [receivers] in in the room will follow the message they received?"

receiver's desk. The sender handed the sheet over to the receiver and then returned to his/her seat. During the delivery process, the experimenter ensured that there would not be any communication between senders and receivers, for instance by talking or through facial expressions. All subjects were informed about the delivery process in the experimental instructions and knew that communication with other subjects implied not being paid their earnings. Once all senders returned to their desk, receivers were asked to type into the computer screen the message they received and to choose one of the ten options to determine the final earnings of each player. Once the experiment ended, subjects were paid in cash. Average earnings were around €15, including a €5 show-up fee.

4. Results

A total of 197 subjects participated in our experiment, 118 as senders and 79 as receivers. Of all the senders, 7 senders switched more than once and 1 sender switched from Message II to Message I as the premium for Message II increased. Since it is not clear what the antisocial premium of these subjects is, we dropped them from the statistical analysis. However, our results are robust to including these subjects and using the number of Message II choices as a measure of antisociality. This leaves us with 39 senders in the *1-Sender* game (19 in Bitter pill and 20 in Deception) and 71 senders in the *2-Sender* game (35 in Bitter pill and 36 in Deception).

4.1. Antisocial premium

Figure 1 plots the cumulative distributions of the senders' antisocial premiums in the *1-Sender* and *2-Sender* games (pooling *Bitter pill* and *Deception* contexts). Consistent with the literature, the figure reveals that many senders are willing to forego large profits in order to act prosocially. More interestingly, having a second sender lowers antisocial premiums since the distribution clearly shifts to the right as one goes from one to two senders.

The difference between the two games is also seen in Figure 2, which depicts the mean antisocial premium depending on the number of senders as well as the context. On average, senders in the *2-Sender* game require €1.40 less for sending Message II than senders in the *1-Sender* game (€1.49 less in *Bitter pill* and €1.33 less in *Deception*). This difference is substantial, considering that the overall mean antisocial premium across both games is only €3.28.

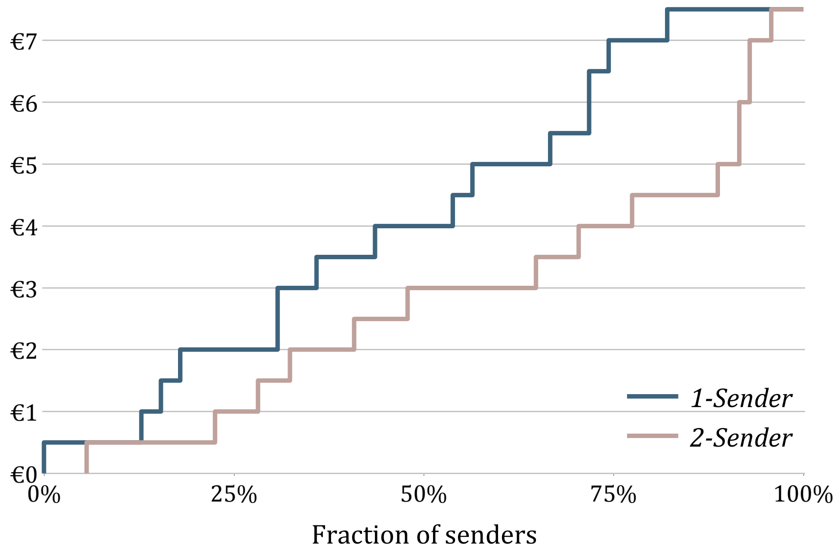


Figure 1. Cumulative distributions of senders' antisocial premiums

In order to evaluate whether these differences are statistically significant, we use interval regressions with the senders' antisocial premium as the dependent variable. This allows us to account for the fact that if a sender switches from Message I to Message II when the latter pays more than €*x*, then we know that her antisocial premium lies in the interval [€*x* – 0.50, €*x*] (see Stewart 1983).[15] All regressions are estimated using robust standard errors and are found in Table A1 of the Appendix. Lastly, since we have a clear directional hypothesis, we report *p*-values of one-tailed tests.

We find that antisocial premiums are significantly lower in the *2-Sender* game compared to the *1-Sender* game ($p = 0.003$ overall; $p = 0.016$ in *Bitter pill*; $p = 0.022$ in *Deception*).[16] We summarize these findings as our first result.

Result 1: *The involvement of a second sender significantly lowers antisocial premiums in both Bitter pill and Deception contexts.*

---

[15] At the extremes, senders who always choose Message I (Message II) have an antisocial premium in the interval [€7.50, ∞) if they played in the *1-Sender* game or as Sender A in the *2-Sender* game or in the interval [€7.00, ∞) if they played as Sender B in the *2-Sender* game. The analogous intervals for senders who always choose Message II are (∞, €0.50] and (∞, €0.00]. These results are robust if we use linear or ordered probit regressions instead of interval regressions.

[16] A difference-in-differences test reveals that the difference between the *1-Sender* and *2-Sender* games does not differ between *Bitter pill* and *Deception* ($p = 0.993$). The difference in antisocial premiums between *Bitter pill* and *Deception* is close to statistical significance in the *1-Sender* game ($p = 0.074$) and is significant in the *2-Sender* game ($p = 0.003$).

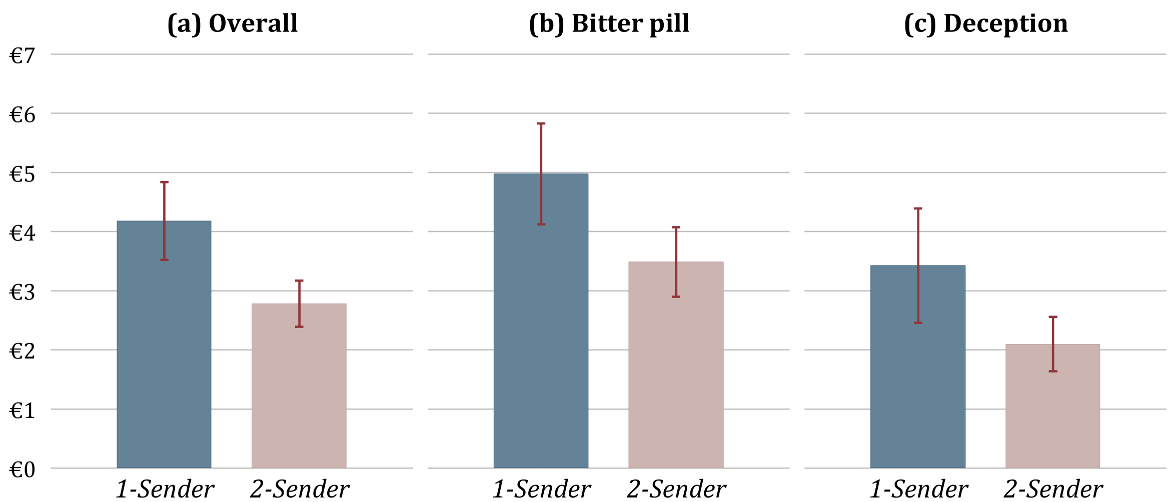


Figure 2. Senders' mean antisocial premium depending on the game and context

Note: Error bars correspond to 95% confidence intervals.

4.2. Normative views and second-order beliefs

Here, we test the hypothesized reasons for why a second sender lowers antisocial premiums.

Normative views

First, we examine the subjects' normative views, i.e., their ratings of how acceptable sending Message II is. Figure 3(a) presents the senders' and receivers' mean acceptability ratings of sending Message II and the senders' mean belief of the receivers' acceptability ratings. More detailed summary statistics are reported in Table 3 as well as in Table A2 in the Appendix. Given that normative views are discrete, ranging from very unacceptable (1) to very acceptable (5) (see Section 3.3), we use ordered probit regressions to test whether differences between games are statistically significant. The regression coefficients are provided in Table A3 of the Appendix. As before, since there is a directional hypothesis, we report *p*-values of one-tailed tests.[17]

On average, senders in the *2-Sender* games think it is significantly more acceptable to send Message II than senders in the *1-Sender* games (2.97 vs. 2.33, *p* = 0.004). Interestingly, the receivers'

[17] Since the senders' normative views were elicited after the message delivery, we cluster standard errors in the *2-Sender* game on the matched pairs. Results are robust to the use of linear regressions instead of ordered probit regressions.

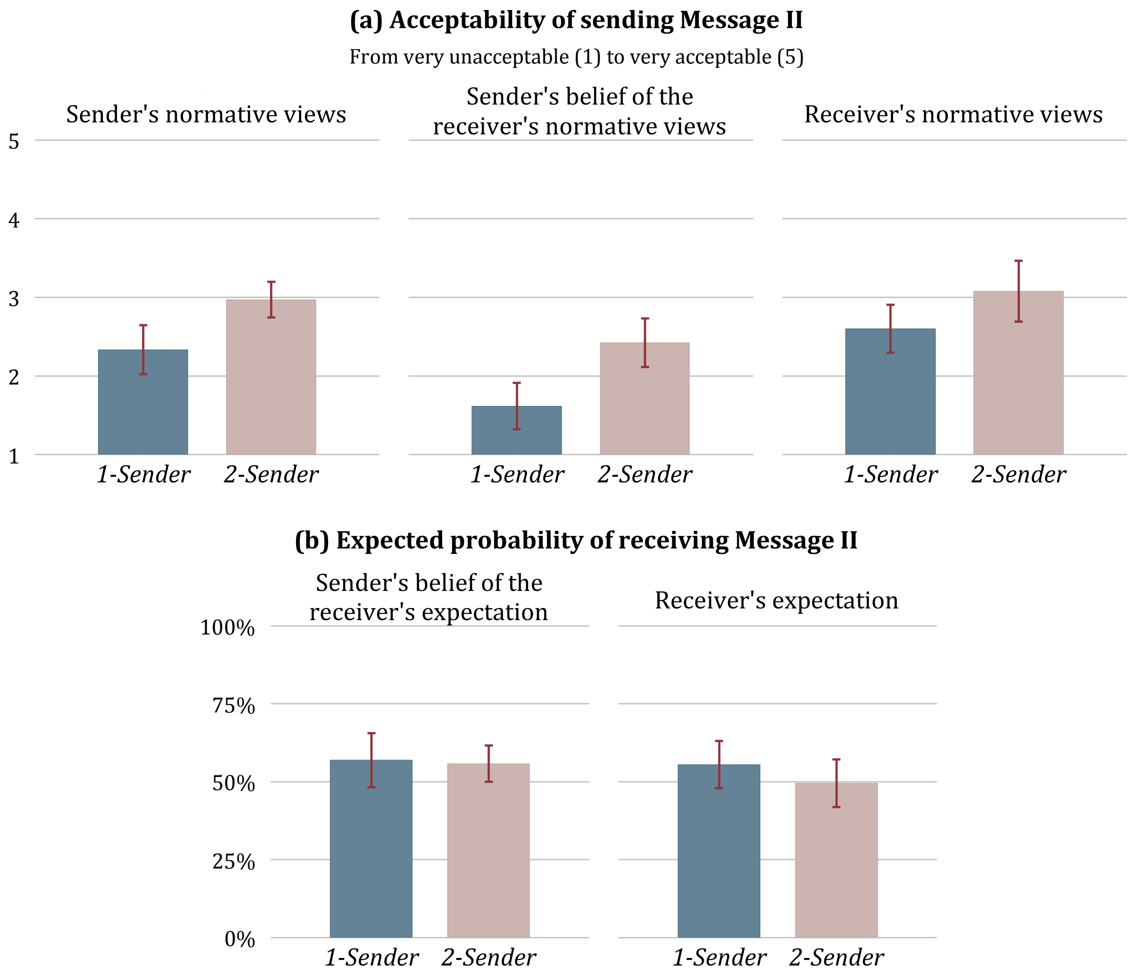


Figure 3. Subjects' mean normative views and beliefs

Note: Error bars correspond to 95% confidence intervals.

acceptability ratings are also higher in the *2-Sender* games (3.08 vs. 2.60, p = 0.051) as are the senders' beliefs of the receivers' acceptability ratings (2.42 vs. 1.62, p=0.004).[18]

A common concern with self-reported measures, such as the subjects' normative views, is that they could be self-serving and depend on their experience in the game. Here we present three pieces

[18] This pattern persists when we examine each context separately. In *Bitter pill*, the difference in acceptability ratings between the *1-Sender* and the *2-Sender* games is –0.69 for senders (p = 0.011), –0.69 for the senders' belief of the receivers' acceptability ratings (p = 0.063), and –0.35 for receivers (p = 0.190). In *Deception*, the difference between the *1-Sender* and the *2-Sender* games is –0.59 for senders (p = 0.050), –0.92 for the senders' beliefs of the receivers' acceptability ratings (p = 0.010) and –0.63 for receivers (p = 0.058).

of evidence suggesting that this is not the case for our elicited normative views. First, antisocial messages are rated more acceptable in *2-Sender* games than in *1-Sender* games by not only senders, but also by receivers. This suggests that the change in normative views is not due to self-serving self-reporting by senders.[19] Second, there is no evidence that senders in the *1-Sender* game rate the acceptability of Message I differently from senders in the *2-Sender* game ($p = 0.145$), which one would expect if normative evaluations are affected by the subjects' experience in the game. Third, we also ran the regressions reported above controlling for subjects' experience up to the point where they reported their normative views. Specifically, we controlled for (i) which one of the two messages was actually delivered, (ii) the choice of the other sender in the *2-Sender* games, and (iii) the senders' earnings if the message is followed by the receiver. We find, once again, that the senders' normative views as well as their beliefs about the receivers' normative views significantly differ between the *1-Sender* and *2-Sender* games ($p = 0.009$ and $p = 0.034$ respectively). Moreover, the control variables are neither jointly significant ($p > 0.770$) nor individually significant ($p > 0.427$) in either regression (see Table A4 in the Appendix for details).

Second-order beliefs

Next, we turn to senders' belief of the receiver's expected probability of receiving Message II, to which we refer to as the senders' *second-order belief*. The senders' second-order belief and the receivers' actual expected probability of receiving Message II are depicted in Figure 3(b) (see Table 3 for more detailed summary statistics). We use Tobit regressions to test for differences across games as belief responses are censored at 0% and 100%. The regression coefficients are provided in Table A5 of the Appendix.[20]

On average, senders think that receivers expect to receive Message II with probability 0.57 in the *1-Sender* games and 0.56 in the *2-Sender* games ($p = 0.628$). The senders' beliefs are fairly accurate as receivers expect to receive Message II with probability 0.56 in the *1-Sender* game and 0.50 in the *2-Sender* game ($p = 0.835$). Hence, contrary to what we initially hypothesized, we do not find evidence

[19] We find that normative views of receivers do not differ significantly from those of senders in either the *1-Sender* ($p = 0.223$) or the *2-Sender* games ($p = 0.727$).

[20] As before, we cluster standard errors on the matched sender pairs in the *2-Sender* game and report *p*-values of one-tailed tests. Results are robust to the use of linear regressions instead of tobit regressions.

Table 3. Means and standard deviations of subjects' normative views and beliefs

	Overall		***Bitter pill***		***Deception***	
	1-Sender	*2-Sender*	*1-Sender*	*2-Sender*	*1-Sender*	*2-Sender*
Senders' normative views of sending Message II	2.33	2.97	2.11	2.80	2.55	3.14
	(1.15)	(1.15)	(1.05)	(1.08)	(1.23)	(1.20)
Senders' belief of the receivers' normative views of receiving Message II	1.62	2.42	1.74	2.43	1.50	2.42
	(1.09)	(1.56)	(1.28)	(1.54)	(0.89)	(1.61)
Receivers' normative views of receiving Message II	2.60	3.08	2.30	2.65	2.90	3.53
	(1.15)	(1.44)	(0.92)	(1.53)	(1.29)	(1.22)
Senders' belief of the receivers' expected probability of receiving Message II	0.57	0.56	0.52	0.49	0.62	0.63
	(0.32)	(0.29)	(0.31)	(0.28)	(0.33)	(0.30)
Receivers' expected probability of receiving Message II	0.56	0.50	0.57	0.34	0.55	0.66
	(0.29)	(0.28)	(0.28)	(0.26)	(0.29)	(0.20)

that senders' second-order beliefs are affected by having a second sender.[21] These findings establish our second result.

Result 2: *The involvement of a second sender makes sending Message II more normatively acceptable. By contrast, the senders' belief of the receivers' expected probability of receiving Message II remain unchanged.*

4.3. Determinants of the antisocial premium

To further understand the determinants of antisocial premiums, we conduct a series of interval regressions of the senders' antisocial premiums, reported in Table 4. The specification in Column I includes the senders' normative views, their second-order beliefs, and a dummy variable indicating whether it is the *Deception* or *Bitter Pill* context. Specification II adds the interaction term of senders' normative views with their second-order beliefs. Finally, specification III includes a set of control variables: (i) the sender's expected probability that the receiver will implement the option mentioned in the message, (ii) the sender's gender, (iii) age, (iv) age squared, and (v) whether the

[21] If we look at each context separately, we find that, compared to the *1-Sender* game, second-order beliefs in the *2-Sender* game are significantly higher in neither *Bitter pill* (p = 0.683) nor *Deception* (p = 0.513).

Table 4. Determinants of the antisocial premium

Note: Interval regressions of senders' antisocial premiums. 'Normative views' are the senders' appropriateness ratings of sending Message II; 'Second-order beliefs' are the senders' belief of the receivers' expected probability of receiving Message II; 'Deception context' is a dummy variable indicating the *Deception* context. Robust standard errors clustered on matched pairs of senders are presented in parentheses. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10 using two-tailed tests.

	***1-Sender* game**			***2-Sender* game**		
Specification	I	II	III	I	II	III
Normative views	−0.39	−0.98*	−1.03*	−0.20	−1.06***	−1.00**
	(0.33)	(0.50)	(0.55)	(0.25)	(0.39)	(0.39)
Second-order belief	−6.32***	−9.15***	−9.39***	−3.01***	−7.25***	−7.50***
	(1.48)	(2.68)	(3.02)	(0.91)	(1.80)	(1.83)
Normative views × Second-order belief		1.13	1.20		1.44***	1.31**
		(0.93)	(1.01)		(0.53)	(0.53)
Deception context	−1.00	−0.94	−1.04	−1.08**	−0.89**	−0.68*
	(0.76)	(0.76)	(0.77)	(0.44)	(0.39)	(0.36)
Constant	9.02***	10.48***	10.76***	5.26***	7.66***	10.29***
	(1.08)	(1.47)	(1.77)	(0.97)	(1.38)	(1.62)
Additional controls	No	No	Yes	No	No	Yes
Observations	39	39	39	71	71	71
χ^2	25.49	30.42	40.57	23.61	34.20	64.74

sender was sender B in the *2-Sender* game.[22] In all regressions, we cluster standard errors on matched sender pairs in the *2-sender* game.

We see a similar pattern across the *1-Sender* and the *2-Sender* games in Table 4. First, normative views have a negative effect on antisocial premiums. That is to say, the premium senders require to send Message II is lower the more acceptable they perceive sending Message II is. Second, the negative coefficients on second-order beliefs indicates that senders are more willing to send the antisocial message if they believe that the receiver expects to receive that message. Third, the interaction between normative views and second-order beliefs is positive, although it is statistically significant only in the *2-Sender* games. In other words, second-order beliefs have a stronger effect on the behavior of senders who think sending Message II is unacceptable compared to senders who

[22] We standardized the control variables so that the constant is comparable across specifications II and III. The coefficients and standard errors of these control variables are presented in Table A6 in the Appendix.

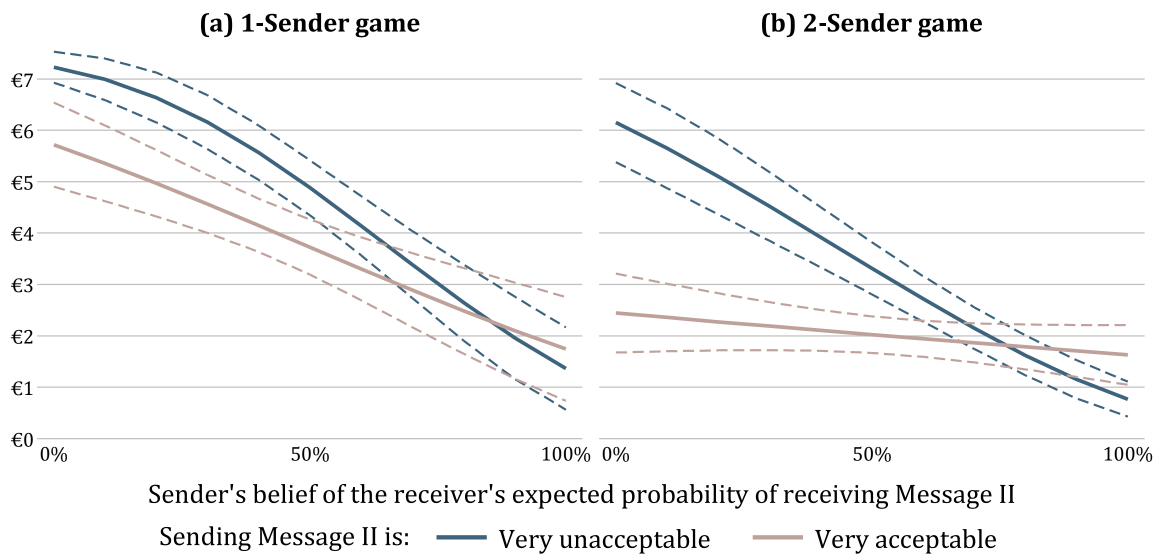


Figure 4. Estimated antisocial premium depending on the sender's normative views and second-order belief

Note: Estimates based on specification III in Table 4. Very unacceptable (acceptable) corresponds to the views of the sender in the 10th (90th) percentile of the normative views distribution. Dotted lines correspond to ± one standard error.

think that sending Message II is acceptable. This general pattern is consistent with models of guilt aversion (Battigalli and Dufwenberg 2007), which predict that senders avoid disappointing the receiver only when their guilt-sensitivity is high.[23]

Despite the similar general pattern, there is a noticeable difference between the *1-Sender* and *2-Sender* games. To visualize this difference, in Figure 4 we use the coefficients of specification III to plot the estimated mean antisocial premium depending on the senders' second-order belief at two different values for their normative views (keeping all other variables at their mean). The dark blue line shows the relationship between antisocial premiums and second-order beliefs for senders who think sending Message II is "very unacceptable" (10th percentile of the distribution of normative views), while the light brown line depicts the same relationship for senders who think sending Message II is "very acceptable" (90th percentile of the distribution of normative views).

The key difference between the *1-Sender* and *2-Sender* games lies in the slopes of the light brown lines. Senders in the *1-Sender* game who think that it is acceptable to send Message II nonetheless

[23] We also conducted regressions substituting the senders' normative views with their belief about the receivers' normative views. The results are presented in Table A7 in the Appendix. We find that the general pattern is similar to the one reported above. However, the interaction between normative views and second-order beliefs is even weaker in the *1-Sender* game.

increase their antisocial premium to avoid disappointing the receiver. By contrast, senders with these normative views in the *2-Sender* game appear to ignore the receiver's expectations.[24] Hence, even though the introduction of a second sender does not impact mean second-order beliefs (see Result 2), it does change the relationship between the second-order beliefs of some senders and their antisocial premiums. This is stated as our third result.

Result 3: *Normative views and second-order beliefs are both important determinants of the senders' antisocial premiums. The relationship between second-order beliefs and antisocial premiums weakens between the 1-Sender and 2-Sender games for senders who view sending Message II as acceptable.*

Senders' emotions

Next, we briefly analyze the senders' emotional reaction. This analysis can be used to corroborate that the senders' hedonic experience is consistent with their behavior across the two games.[25] Our analysis focuses on the amount of guilt that senders experience the moment they see the final outcome of the game (measured in a 1 to 7 Likert scale) depending on whether they sent the antisocial or the prosocial message.[26]

On average, senders experience significantly more guilt after sending the antisocial Message II than after sending the prosocial Message I. The difference is substantial: 4.23 vs. 1.09 in the *1-Sender* game and 2.68 vs. 1.62 in the *2-Sender* game ($p < 0.026$). More tellingly, we also find that, on average, senders experience significantly less guilt after sending Message II in the *2-Sender* game compared

[24] When testing whether the estimated social premiums differ between games, we find that senders who think that it is acceptable to send Message II have significantly higher antisocial premiums in the *1-Sender* game as long as their second-order belief is less than 0.59 (Wald tests, $p < 0.05$).

[25] Summary statistics for all emotions are provided in Table A8 in the Appendix. Moreover, although our interest is on guilt, we also performed a similar analysis for the other elicited emotions. Table A10 contains the regression results of other negative emotions (shame and anger) and Table A11 of positive emotions (happiness and gratitude). By and large, these results are in line with the results for guilt.

[26] In the *1-Sender* game, 22 Message I's and 17 Message II's were delivered; while in the *2-Sender* game, 27 Message I's and 11 Message II's were delivered. From these, we drop the six instances where the final outcome was not a direct consequence of the senders' choices because the receiver chose a different option from the one suggested in the message. However, our results remain unchanged if we include these observations.

to the *1-Sender* game (2.68 vs. 4.23, $p < 0.006$).[27] Hence, the effect of having a second sender is not only evident in the senders' behavior but also in their emotional state.[28] This leads us to our fourth result.

Result 4: *Senders experience less guilt for sending Message II if a second sender is involved.*

4.4. Diffusion of responsibility and active participation

An important question arising from our main findings is: does diffusion of responsibility require sender B's active participation in the decision making or is the presence of a second sender enough to reduce antisocial premiums? To answer this question, we ran additional sessions using a variation of the *2-Sender* game that we refer to as the *Passive-Sender-B* game.

In the *Passive-Sender-B* game, sender B is present, delivers the message, and receives the same payoffs as in the *2-Sender* game, but sender B does not have any say on the content of the message that is sent to the receiver. The message is picked solely by sender A using the same procedure as in the *1-Sender* game. Detailed instructions are provided in the Appendix. We ran two sessions of the *Passive-Sender-B* game in the *Deception* context.[29] A total of 66 subjects participated in these sessions, 22 sender A's, 22 sender B's, and 22 receivers.

As mentioned in Section 3, some prominent outcome-based models of social preferences (e.g., Fehr and Schmidt 1999; Bolton and Ockenfels 2000; Charness and Rabin 2002) predict that adding a

---

[27] These *p*-values are obtained from linear regressions of the senders' experienced guilt reported in Table A12 of the Appendix. In addition, in Table A9 of the Appendix, we present regressions that analyze the association between the senders' guilt and their normative views whilst controlling for the context and the senders' individual characteristics. We find that senders who deliver Message II experience more guilt the more they consider that sending Message II is normatively unacceptable. Interestingly, this effect is much stronger in the *1-Sender* game compared to the *2-Sender* game.

[28] A possible concern may be that these differences are the consequence of emotions being self-reported. For instance, senders might not report their true emotions and instead report the emotional reaction they think the experimenter expects. We think that this is an unlikely explanation for the difference between games. That is, it is hard to see how subjects could anticipate that the `expected' emotional reaction is more guilt for Message II in the 1-Sender than in the 2-Sender game when subjects took part in only one of these games.

[29] We decided against running sessions of the *Passive-Sender-B* game in the *Bitter pill* context because the introduction of a second sender lowers anti-social premiums similarly in both contexts. This way, we could concentrate our resources on one context (*Deception*) and increase the power of the comparison between the three games.

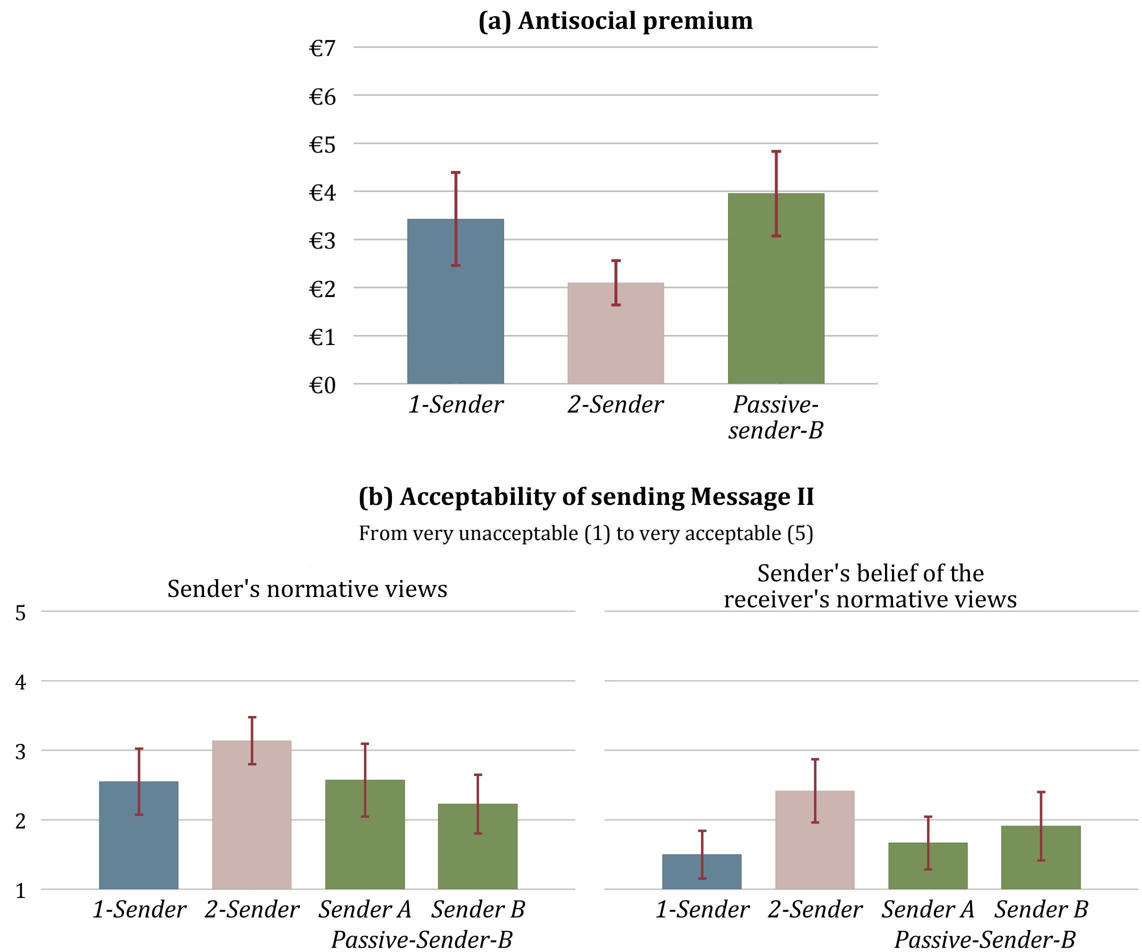


Figure 5. Subjects' mean antisocial premiums and normative views including those of Sender A and Sender B from the *Passive-Sender-B* game

Note: The *1-Sender* and *2-Sender* games include subjects from only the deception context. Error bars correspond to 95% confidence intervals.

second sender lowers antisocial premiums. This prediction, however, holds irrespective of whether the second sender takes an active part in the decision-making process or not. Hence, with the *Passive-Sender-B* game, we are able to determine the extent to which these models explain the observed reduction in antisocial premiums between the *1-Sender* and *2-Sender* games.

Figure 5(a) depicts the mean antisocial premium in the *Passive-Sender-B* game as well as in the *Deception* context of the *1-Sender* and *2-Sender* games. Like in section 4.1, we use interval regressions to evaluate whether differences are statistically significant (available in Table A1 of the Appendix). We find that antisocial premiums in the *Passive-Sender-B* game are close to those in the *1-Sender* game (€3.95 vs. €3.43 on average; $p = 0.558$) and significantly higher than antisocial premiums in the *2-Sender* game (€3.95 vs. €2.10 on average; $p = 0.004$). Therefore, we conclude that active participation of a partner in crime is necessary for diffusion of responsibility to occur.

A similar pattern is observed when we compare normative views across the three games. Figure 5(b) depicts the senders' mean acceptability ratings of sending Message II and their belief of the receivers' acceptability ratings in the *1-Sender* and *2-Sender* games as well as in the *Passive-Sender-B* game for both the active Sender A and the passive Sender B. By and large, we see that the senders' acceptability ratings of the active and passive senders in the *Passive-Sender-B* game are similar to those of senders in the *1-Sender* game and below those of senders in the *2-Sender* game.[30] In the Appendix, we further demonstrate that the beliefs and emotions of senders in the *Passive-Sender-B* game follow the same pattern as those of senders in the *1-Sender* game and display similar differences vis-à-vis senders in the *2-Sender* game. Namely, there are no noticeable differences in second-order beliefs and lower levels of experienced guilt if the receiver follows Message II in the *Passive-Sender-B* game compared to the *1-Sender* game (for details see Tables A1, A2, A3, A6, A7, A8 and A12). We summarize our findings as our last result.

Result 5: *The addition of a passive second sender does not lower antisocial premiums compared to the game with one sender. Consistent with this result, a passive second sender does not make sending Message II more normatively acceptable.*

---

[30] Like in section 4.2, we use ordered probit regressions to evaluate whether differences are statistically significant (available in Table A4 of the Appendix). There are no statistical differences between senders in the *1-Sender* game and senders A in the *Passive-Sender-B* game (for acceptability ratings, 2.55 vs. 2.57, $p = 0.979$; for belief of the receivers' acceptability ratings 1.50 vs. 1.67, $p = 0.707$) or senders B in the *Passive-Sender-B* game (for acceptability ratings, 2.55 vs. 2.23, $p = 0.365$; for belief of the receivers' acceptability ratings 1.50 vs. 1.91, $p = 0.707$). By contrast, senders in the *2-Sender* game tend to view sending Message II to be more acceptable than senders A in the *Passive-Sender-B* game (for acceptability ratings, 3.14 vs. 2.57, $p = 0.117$; for belief of the receivers' acceptability ratings 2.42 vs. 1.67, $p = 0.055$), and senders B in the *Passive-Sender-B* game (for acceptability ratings, 3.14 vs. 2.23, $p = 0.009$; for belief of the receivers' acceptability ratings 2.42 vs. 1.91, $p = 0.227$). Similarly, the acceptability ratings of receivers in the *Passive-Sender-B* game are similar to those of receivers in the *1-Sender* game (2.73 vs. 2.90, $p = 0.337$) but not in the *2-Sender* game (2.73 vs. 3.53, $p = 0.048$).

5. Conclusion

In this study, we first present evidence that individuals tend to behave more antisocially with a partner than when acting alone. We attribute this phenomenon to diffusion of responsibility, as our games are designed to rule out other popular explanations for this phenomenon. In the *2-Sender* game, senders submit their decisions simultaneously via a choice list so that joint decisions are made without allowing any type of interaction. The absence of interaction between senders allows us to eliminate explanations such as peer effects and information exchange through the deliberation process, while the lack of a typical market setting rules out explanations such as social information revealed through the bids and asks of others and the introduction of a more materialistic framing.

Moreover, our interpretation of diffusion of responsibility is also consistent with the senders' normative views and their emotional reactions. We find that senders think that sending antisocial messages is more acceptable in the presence of a second sender. Interestingly, this difference in normative views in also held by receivers, which suggests that diffusion of responsibility is not a self-serving shift in the senders' other-regarding concerns. In other words, senders are not using the presence of the second sender as an "excuse" to behave badly. Instead, it is a more general shift in how unacceptable the antisocial option is perceived.

Finally, results from our *Passive-Sender-B* game, in which the second sender does not take on an active role, suggests that the key driver for diffusion of responsibility is the partner's active involvement in the decision process, and is not the senders' preferences over their partner's final earnings. This finding is inconsistent with outcome-based models of social preferences as an explanation for diffusion of responsibility.

References

Adolphs, Ralph. 2002. "Recognizing Emotion from Facial Expressions: Psychological and Neurological Mechanisms." *Behavioral and Cognitive Neuroscience Reviews* 1 (1): 21–62. doi:10.1177/1534582302001001003.

Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *The Economic Journal* 100 (401): 464. doi:10.2307/2234133.

Barr, Abigail, and Georgia Michailidou. 2017. "Complicity without Connection or Communication." *Journal of Economic Behavior & Organization* 142 (October): 1–10. doi:10.1016/j.jebo.2017.07.013.

Bartling, Björn, and Urs Fischbacher. 2012. "Shifting the Blame: On Delegation and Responsibility." *The Review of Economic Studies* 79 (1): 67–87. doi:10.1093/restud/rdr023.

Bartling, Björn, Urs Fischbacher, and Simeon Schudy. 2015. "Pivotality and Responsibility Attribution in Sequential Voting." *Journal of Public Economics* 128 (August): 133–39. doi:10.1016/j.jpubeco.2015.03.010.

Bartling, Björn, Roberto A. Weber, and Lan Yao. 2015. "Do Markets Erode Social Responsibility?" *The Quarterly Journal of Economics* 130 (1): 219–66. doi:10.1093/qje/qju031.

Battigalli, Pierpaolo, Gary Charness, and Martin Dufwenberg. 2013. "Deception: The Role of Guilt." *Journal of Economic Behavior & Organization* 93 (September): 227–32. doi:10.1016/j.jebo.2013.03.033.

Battigalli, Pierpaolo, and Martin Dufwenberg. 2007. "Guilt in Games." *American Economic Review* 97 (2): 170–76. doi:10.1257/aer.97.2.170.

Ben-Shakhar, Gershon, Gary Bornstein, Astrid Hopfensitz, and Frans van Winden. 2007. "Reciprocity and Emotions in Bargaining Using Physiological and Self-Report Measures." *Journal of Economic Psychology* 28: 314–23.

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96 (5): 1652–78. doi:10.1257/aer.96.5.1652.

Bicchieri, Cristina. 2017. *Norms in the Wild*. Oxford University Press. doi:10.1093/acprof:oso/9780190622046.001.0001.

Bicchieri, Cristina, and Erte Xiao. 2009. "Do the Right Thing: But Only If Others Do so." *Journal of Behavioral Decision Making* 22 (2): 191–208. doi:10.1002/bdm.621.

Bolton, Gary E, and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review* 90 (1): 166–93. doi:10.1257/aer.90.1.166.

Bornstein, Gary, and Ilan Yaniv. 1998. "Individual and Group Behavior in the Ultimatum Game: Are Groups More 'rational' players?" *Experimental Economics* 1 (1): 101–8. doi:10.1007/BF01426217.

Bradley, Margaret M, and Peter J Lang. 2000. "Measuring Emotion: Behavior, Feeling and Physiology." In *Cognitive Neuroscience of Emotions*, edited by Richard D. Lang and Lynn Nadel, 242–76. Oxford: Oxford University Press.

Cason, Timothy N, and Vai-Lam Mui. 1997. "A Laboratory Study of Group Polarisation in the Team Dictator Game." *The Economic Journal* 107 (444): 1465–83. doi:10.1111/j.1468-0297.1997.tb00058.x.

Charness, Gary, and Mathew Rabin. 2002. "Understanding Social Preferences with Simple Tests." *The Quarterly Journal of Economics* 117 (3): 817–69. doi:10.1162/003355302760193904.

Chytilová, Julie, and Václav Korbel. 2014. “Individual and Group Cheating Behavior: A Field Experiment with Adolescents.” IES Working Paper 06/2014.

Coffman, Lucas C. 2011. “Intermediation Reduces Punishment (and Reward).” *American Economic Journal: Microeconomics* 3 (4): 77–106. doi:10.1257/mic.3.4.77.

Cohen, Taya R, Brian C Gunia, Sun Young Kim-Jun, and J Keith Murnighan. 2009. “Do Groups Lie More than Individuals? Honesty and Deception as a Function of Strategic Self-Interest.” *Journal of Experimental Social Psychology* 45 (6): 1321–24. doi:10.1016/j.jesp.2009.08.007.

Conrads, Julian, Bernd Irlenbusch, Rainer Michael Rilke, and Gari Walkowitz. 2013. “Lying and Team Incentives.” *Journal of Economic Psychology* 34 (February): 1–7. doi:10.1016/j.joep.2012.10.011.

Cox, James C. 2002. “Trust, Reciprocity, and Other-Regarding Preferences: Groups Vs. Individuals and Males Vs. Females.” In *Experimental Business Research*, edited by Rami Zwick and Amnon Rapoport, 331–50. Boston, MA: Springer US. doi:10.1007/978-1-4757-5196-3_14.

Dana, Jason, Roberto A. Weber, and Jason Xi Kuang. 2007. “Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness.” *Economic Theory* 33 (1). Springer-Verlag: 67–80. doi:10.1007/s00199-006-0153-z.

Danilov, Anastasia, Torsten Biemann, Thorn Kring, and Dirk Sliwka. 2013. “The Dark Side of Team Incentives: Experimental Evidence on Advice Quality from Financial Service Professionals.” *Journal of Economic Behavior & Organization* 93: 266–72. doi:10.1016/j.jebo.2013.03.012.

Deckers, Thomas, Armin Falk, Fabian Kosse, and Nora Szech. 2016. “Homo Moralis: Personal Characteristics, Institutions, and Moral Decision-Making.” Working Paper 5800.

Ellingsen, Tore, and Magnus Johannesson. 2004. “Promises, Threats and Fairness.” *The Economic Journal* 114 (495): 397–420. doi:10.1111/j.1468-0297.2004.00214.x.

Engl, Florian. 2017. “A Theory of Causal Responsibility Attribution.” Working Paper 2932769.

Erkut, Hande, Daniele Nosenzo, and Martin Sefton. 2015. “Identifying Social Norms Using Coordination Games: Spectators vs. Stakeholders.” *Economics Letters* 130 (May): 28–31. doi:10.1016/j.econlet.2015.02.021.

Falk, Armin, and Nora Szech. 2013. “Morals and Markets.” *Science* 340 (6133). American Association for the Advancement of Science: 707–11. doi:10.1126/science.1231566.

———. 2016. “Diffusion of Being Pivotal and Immoral Outcomes.” Working Paper 2016-013.

Fehr, Ernst, and Klaus M Schmidt. 1999. “A Theory of Fairness, Competition, and Cooperation.” *The Quarterly Journal of Economics* 114 (3): 817–68. doi:10.1162/003355399556151.

———. 2006. “The Economics of Fairness, Reciprocity and Altruism-Experimental Evidence and New Theories.” In *Handbook on the Economics of Giving, Reciprocity and Altruism*, edited by Serge C Kolm and Jean M Ythier, 615–91. Amsterdam: Elsevier. doi:10.1016/S1574-0714(06)01008-6.

Fischbacher, Urs. 2007. “Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments.” *Experimental Economics* 10 (2): 171–78. doi:10.1007/s10683-006-9159-4.

Fischbacher, Urs, and Franziska Föllmi-Heusi. 2013. “Lies In Disguise-An Experimental Study On Cheating.” *Journal of the European Economic Association* 11 (3): 525–47. doi:10.1111/jeea.12014.

Garofalo, Orsola, and Christina Rott. 2017. “Shifting Blame? Experimental Evidence of Delegating Communication.” *Management Science*, August. doi:10.1287/mnsc.2017.2782.

Gino, Francesca, Shahar Ayal, and Dan Ariely. 2013. “Self-Serving Altruism? The Lure of Unethical Actions That Benefit Others.” *Journal of Economic Behavior & Organization* 93 (September): 285–92. doi:10.1016/j.jebo.2013.04.005.

Gino, Francesca, and Lamar Pierce. 2010. “Lying to Level the Playing Field: Why People May Dishonestly Help or Hurt Others to Create Equity.” *Journal of Business Ethics* 95 (S1). Springer Netherlands: 89–103. doi:10.1007/s10551-011-0792-2.

Gneezy, Uri. 2005. “Deception: The Role of Consequences.” *American Economic Review* 95 (1): 384–94. doi:10.1257/0002828053828662.

Greiner, Ben. 2015. “Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE.” *Journal of the Economic Science Association* 1 (1). Springer US: 114–25. doi:10.1007/s40881-015-0004-4.

Hamman, John R, George Loewenstein, and Roberto A Weber. 2010. “Self-Interest through Delegation: An Additional Rationale for the Principal-Agent Relationship.” *American Economic Review* 100 (4): 1826–46. doi:10.1257/aer.100.4.1826.

Hopfensitz, Astrid, and Ernesto Reuben. 2009. “The Importance of Emotions for the Effectiveness of Social Punishment.” *The Economic Journal*, TI-2005-075/1, 119 (540). Blackwell Publishing: 1534–59. doi:10.1111/j.1468-0297.2009.02288.x.

Keck, Steffen. 2014. “Group Reactions to Dishonesty.” *Organizational Behavior and Human Decision Processes* 124 (1): 1–10. doi:10.1016/j.obhdp.2013.12.006.

Kirchler, Michael, Jürgen Huber, Matthias Stefan, and Matthias Sutter. 2016. “Market Design and Moral Behavior.” *Management Science* 62 (9). INFORMS: 2615–25. doi:10.1287/mnsc.2015.2246.

Kocher, Martin G., Simeon Schudy, and Lisa Spantig. 2017. “I Lie? We Lie! Why? Experimental Evidence on a Dishonesty Shift in Groups.” *Management Science*, August. doi:10.1287/mnsc.2017.2800.

Kocher, Martin G, and Matthias Sutter. 2007. "Individual versus Group Behavior and the Role of the Decision Making Procedure in Gift-Exchange Experiments." *Empirica* 34 (1). Kluwer Academic Publishers-Plenum Publishers: 63–88. doi:10.1007/s10663-006-9026-8.

Krupka, Erin L, and Roberto A Weber. 2013. "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?" *Journal of the European Economic Association* 11 (3): 495–524. doi:10.1111/jeea.12006.

Luhan, Wolfgang J, Martin G Kocher, and Matthias Sutter. 2009. "Group Polarization in the Team Dictator Game Reconsidered." *Experimental Economics* 12 (1). Springer US: 26–41. doi:10.1007/s10683-007-9188-7.

Muehlheusser, Gerd, Andreas Roider, and Niklas Wallmeier. 2015. "Gender Differences in Honesty: Groups versus Individuals." *Economics Letters* 128 (March): 25–29. doi:10.1016/j.econlet.2014.12.019.

Nielsen, Kirby, Puja Bhattacharya, John H. Kagel, and Arjun Sengupta. 2017. "Teams Promise But Do Not Deliver." Working Paper 2998300.

Oexl, Regine, and Zachary J Grossman. 2013. "Shifting the Blame to a Powerless Intermediary." *Experimental Economics* 16 (3). Springer US: 306–12. doi:10.1007/s10683-012-9335-7.

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

Reuben, Ernesto, and Arno Riedl. 2013. "Enforcement of Contribution Norms in Public Good Games with Heterogeneous Populations." *Games and Economic Behavior* 77 (1): 122–37. doi:10.1016/j.geb.2012.10.001.

Reuben, Ernesto, and Frans van Winden. 2010. "Fairness Perceptions and Prosocial Emotions in the Power to Take." *Journal of Economic Psychology* 31 (6). Elsevier B.V.: 908–22. doi:10.1016/j.joep.2010.08.001.

Rothenhäusler, Dominik, Nikolaus Schweizer, and Nora Szech. 2017. "Guilt in Voting and Public Good Games." *European Economic Review*, August. doi:10.1016/j.euroecorev.2017.08.001.

Stewart, Mark B. 1983. "On Least Squares Estimation When the Dependent Variable Is Grouped." *The Review of Economic Studies* 50 (4): 737. doi:10.2307/2297773.

Sutter, Matthias. 2009. "Deception Through Telling the Truth?! Experimental Evidence From Individuals and Teams." *The Economic Journal* 119 (534). Blackwell Publishing Ltd: 47–60. doi:10.1111/j.1468-0297.2008.02205.x.

van Winden, Frans. 2007. "Affect and Fairness in Economics." *Social Justice Research* 20 (1): 35–52.

Weisel, Ori, and Shaul Shalvi. 2015. "The Collaborative Roots of Corruption." *Proceedings of the National Academy of Sciences* 112 (34). National Academy of Sciences: 10651–56. doi:10.1073/pnas.1423035112.

Wiltermuth, Scott S. 2011. “Cheating More When the Spoils Are Split.” *Organizational Behavior and Human Decision Processes* 115 (2): 157–68. doi:10.1016/j.obhdp.2010.10.001.

Appendix for "Partners in Crime: Diffusion of Responsibility in Antisocial Behaviors"

ABSTRACT

This document contains the supplementary materials for the paper Behnk, Hao, and Reuben (2017). It is organized as follows: Section A1 contains a sample of the instructions and screenshots used in the experiment; Section A2 contains the regressions presented in the paper including the coefficients of the control variables and additional statistical analysis reported in the paper but not fully described there due to space constraints.

A1. Instructions and screenshots

This section contains instructions and screenshots used in the *Deception* treatments. The instructions used in the *Bitter pill* treatments are almost identical and are available from the authors upon request.

A1.1. Instructions for the 1-Sender game in Deception

You are participating in a study on economic decision-making. You have already earned **€5** for showing up on time. Please read these instructions carefully as they describe how you can earn **additional** money. You will be paid all your earnings in cash.

Please do not talk or communicate with other participants in any way. If you have questions, raise your hand and one of us will help you.

In the study, all participants are randomly assigned to groups of two. Within each group, the computer randomly assigns participants to the roles of **Player 1** and **Player 2**. You will be informed of your role on the computer screen.

Summary of the study

- There are ten options with payments for each player. Player 1 is informed of the payment each player receives in each option. Player 2 does not receive this information.

- **Player 1 chooses one message out of the two available messages** to be sent to Player 2. Each message states that a specific option is the option that gives the highest payment to Player 2.
- Player 1 delivers the message to Player 2 in person.
- **Player 2 chooses an option** that determines the earnings of both players.

Specific instructions

There are ten options, each one labelled with a unique letter: A, B, C, D, E, F, G, H, I, or J. The computer will randomly assign one option to pay **€10 to Player 1 and €10 to Player 2** and another option to pay **€17 to Player 1 and €3 to Player 2**. The remaining eight options pay **€4 to Player 1 and €0 to Player 2**.

How much each player earns in each option will be shown only to Player 1. The following table is an example of how payments could be assigned to the various options and how this information would be presented to Player 1.

Option	A	B	**C**	D	**E**	F	G	H	I	J
Player 1’s payment	4	4	10	4	17	4	4	4	4	4
Player 2’s payment	0	0	10	0	3	0	0	0	0	0

Player 2 will not know which options provide positive earnings for him/her. The table below shows what Player 2 will see.

Option	A	B	C	D	E	F	G	H	I	J
Player 1’s payment	?	?	?	?	?	?	?	?	?	?
Player 2’s payment	?	?	?	?	?	?	?	?	?	?

The only information that Player 2 receives regarding the payments of the various options is the message chosen and delivered by Player 1. After receiving the message, Player 2 chooses one of the ten options. The option chosen by Player 2 determines the earnings of both players in the group.

Player 1 chooses a message

Player 1 chooses **one message** for Player 2. There are **two available messages**. Each message corresponds to one of the two options with positive earnings for both players.

Message I corresponds to the option that pays €10 to Player 2. The message reads "Option *<letter of option that pays €10 to Player 2>* will earn you 10 euros".

Message II corresponds to the option that pays €3 to Player 2. The message reads "Option *<letter of option that pays €3 to Player 2>* will earn you 10 euros".

Note that Player 1 cannot choose a message that corresponds to an option that pays €0 to Player 2. Therefore, when Player 2 receives a message, he/she will not know whether the option mentioned in the message pays him/her €10 or €3, but he/she can be certain that the option does not pay him/her €0.

Example

Suppose that the computer randomly assigns payments to options as shown in the table below.

Option	A	B	C	**D**	E	**F**	G	H	I	J
Player 1's payment	4	4	4	17	4	10	4	4	4	4
Player 2's payment	0	0	0	3	0	10	0	0	0	0

In this case, Player 2 can receive one of the following two messages:

"Option F will earn you 10 euros"

"Option D will earn you 10 euros"

Player 1 delivers the message to Player 2. If the option mentioned in the message coincides with the option subsequently chosen by Player 2, then **Player 1 pays a cost between €0 and €6.50 for delivery**. The screens below will be used to determine which message Player 1 delivers and how much Player 1 pays. Each screen displays a list containing 14 rows, each row representing a possible cost. Player 1 must decide between Message I and Message II in each of the 14 rows. Specifically, in each row, Player 1 decides between:

- Choosing **Message I** and paying a cost of €0 for delivering this message.
- Choosing **Message II** and paying a cost **specified in that row** for delivering this message.

After Player 1 has made his/her decisions, **one of the 14 rows will be randomly selected** by the computer to determine which message will be delivered to Player 2. All rows have the same chance of being selected; therefore, you should make your decision in each row seriously.

Decisions of Player 1

You are Player 1

Option	A	B	C	D	E	F	G	H	I	J
Player 1's payment	10	4	4	4	4	4	4	17	4	4
Player 2's payment	10	0	0	0	0	0	0	3	0	0

Please decide between Message I and Message II in **each row.**

Message I:
Option A will earn you 10 euros

Player 2 earns **€10** if he/she chooses Option A.

Message II:
Option H will earn you 10 euros

Player 2 earns **€3** if he/she chooses Option H.

Row	Cost	Your payment		Cost	Your payment
1	0	10	○ ○	0.0	17.0
2	0	10	○ ○	0.5	16.5
3	0	10	○ ○	1.0	16.0
4	0	10	○ ○	1.5	15.5
5	0	10	○ ○	2.0	15.0
6	0	10	○ ○	2.5	14.5
7	0	10	○ ○	3.0	14.0
8	0	10	○ ○	3.5	13.5
9	0	10	○ ○	4.0	13.0
10	0	10	○ ○	4.5	12.5
11	0	10	○ ○	5.0	12.0
12	0	10	○ ○	5.5	11.5
13	0	10	○ ○	6.0	11.0
14	0	10	○ ○	6.5	10.5

Player 1 will deliver the message determined by his/her choices in the selected row in the following way:

- In the selected row, if **Player 1 chose Message I**, then **Player 1 delivers Message I**. In this case, if Player 2 chooses the option corresponding to Message I, then both Player 1 and Player 2 earn €10.
- In the selected row, if **Player 1 chose Message II**, then **Player 1 delivers Message II**. In this case, if Player 2 chooses the option corresponding to Message II, then Player 1 earns €17 minus the cost specified in that row and Player 2 earns €3.

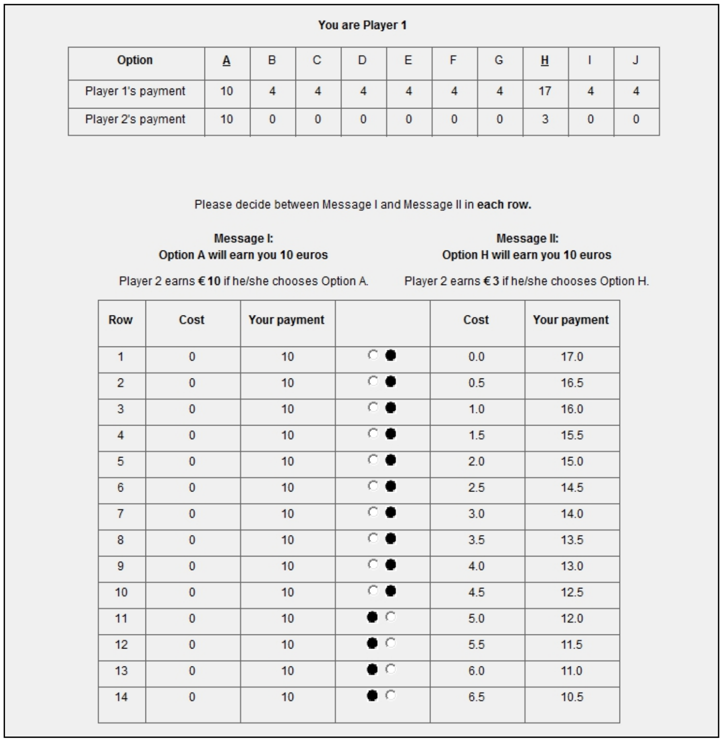
Player 2 will **not** be informed which row was selected by the computer.

Examples

Suppose that Player 1 makes the choices shown below. In this example, Player 1 is willing to pay at maximum €4.5 for delivering Message II. Given these choices, the following occurs if the computer randomly selects one of the rows below:

- **Row 9:** Since Player 1 chose Message II then he/she delivers Message II. Thereafter, if Player 2 chooses the option corresponding to Message II, then Player 1 earns €17 – €4 = €13 and Player 2 earns €3.
- **Row 12:** Since Player 1 chose Message I then he/she delivers Message I. Thereafter, if Player 2 chooses the option corresponding to Message I, then both Player 1 and Player 2 earn €10.

Decisions of Player 1



You are Player 1

Option	A	B	C	D	E	F	G	H	I	J
Player 1's payment	10	4	4	4	4	4	4	17	4	4
Player 2's payment	10	0	0	0	0	0	0	3	0	0

Please decide between Message I and Message II in **each row.**

Message I:
Option A will earn you 10 euros
Player 2 earns **€ 10** if he/she chooses Option A.

Message II:
Option H will earn you 10 euros
Player 2 earns **€ 3** if he/she chooses Option H.

Row	Cost	Your payment		Cost	Your payment
1	0	10	○ ●	0.0	17.0
2	0	10	○ ●	0.5	16.5
3	0	10	○ ●	1.0	16.0
4	0	10	○ ●	1.5	15.5
5	0	10	○ ●	2.0	15.0
6	0	10	○ ●	2.5	14.5
7	0	10	○ ●	3.0	14.0
8	0	10	○ ●	3.5	13.5
9	0	10	○ ●	4.0	13.0
10	0	10	○ ●	4.5	12.5
11	0	10	● ○	5.0	12.0
12	0	10	● ○	5.5	11.5
13	0	10	● ○	6.0	11.0
14	0	10	● ○	6.5	10.5

Suppose that Player 1 makes the choices shown below. In this example, Player 1 is willing to pay at maximum €2 for delivering Message II. Given these choices, the following occurs if the computer randomly selects one of the rows below:

- **Row 4:** Since Player 1 chose Message II then he/she delivers Message II. Thereafter, if Player 2 chooses the option corresponding to Message II, then Player 1 earns €17 – €1.5 = €15.5 and Player 2 earns €3.
- **Row 6:** Since Player 1 chose Message I then he/she delivers Message I. Thereafter, if Player 2 chooses the option corresponding to Message I, then both Player 1 and Player 2 earn €10.

Decisions of Player 1

You are Player 1

Option	A	B	C	D	E	F	G	H	I	J
Player 1's payment	10	4	4	4	4	4	4	17	4	4
Player 2's payment	10	0	0	0	0	0	0	3	0	0

Please decide between Message I and Message II in **each row.**

Message I:
Option A will earn you 10 euros

Player 2 earns **€ 10** if he/she chooses Option A.

Message II:
Option H will earn you 10 euros

Player 2 earns **€ 3** if he/she chooses Option H.

Row	Cost	Your payment		Cost	Your payment
1	0	10	● ○	0.0	17.0
2	0	10	● ○	0.5	16.5
3	0	10	● ○	1.0	16.0
4	0	10	● ○	1.5	15.5
5	0	10	● ○	2.0	15.0
6	0	10	○ ●	2.5	14.5
7	0	10	○ ●	3.0	14.0
8	0	10	○ ●	3.5	13.5
9	0	10	○ ●	4.0	13.0
10	0	10	○ ●	4.5	12.5
11	0	10	○ ●	5.0	12.0
12	0	10	○ ●	5.5	11.5
13	0	10	○ ●	6.0	11.0
14	0	10	○ ●	6.5	10.5

Player 1 delivers the message to Player 2 in person

Once the message is determined, Player 1 will see a screen like the one below.

Please **write down the following message** on the sheet of paper located on your desk and **wait until an experimenter arrives.**

The experimenter will check that the message you wrote down coincides with the message below.

Option J will earn you 10 euros

To deliver the message, Player 1 will first **write down the message on the sheet of paper** located on his/her desk. Then, Player 1 will wait until an experimenter arrives. The experimenter will check whether the message written on the sheet of paper is identical to the message shown on the screen.

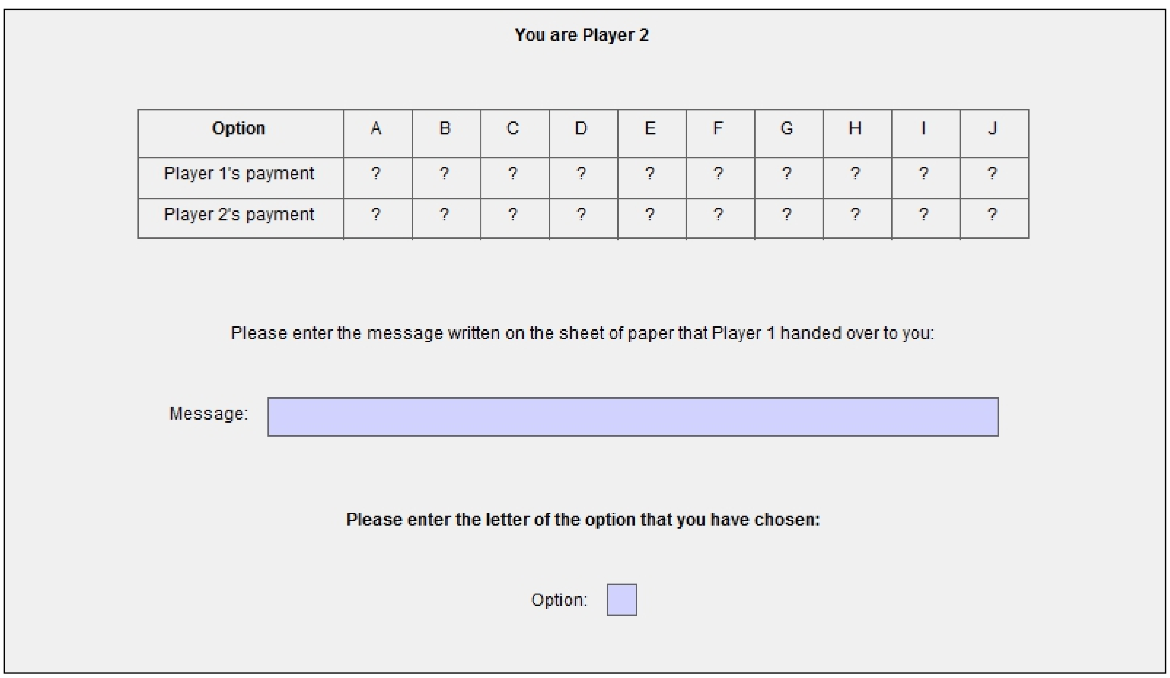
Note that, like Player 2, the experimenter will not know to which payment the option in the message corresponds.

The experimenter will then walk with Player 1 to the desk of the Player 2 of his/her group. At this point, Player 1 will hand the paper with the message to Player 2 and then walk back to his/her desk.

Remember that **any kind of communication between the players is prohibited**, including gestures and facial expressions. In addition, Player 1 is not allowed to write down anything else other than the message on the sheet of paper. Any participant who does not comply with these rules will **not be paid** at the end of the study.

Player 2 chooses an option

Player 2 knows that there are two options with positive payments for him/her, but he/she does not know which two of the ten options contain these payments. **The only information that Player 2 receives is the message delivered to him/her by Player 1.** After receiving the message, Player 2 sees a screen like this:


You are Player 2

Option	A	B	C	D	E	F	G	H	I	J
Player 1's payment	?	?	?	?	?	?	?	?	?	?
Player 2's payment	?	?	?	?	?	?	?	?	?	?

Please enter the message written on the sheet of paper that Player 1 handed over to you:

Message:

Please enter the letter of the option that you have chosen:

Option:

On this screen, Player 2 first confirms the message he/she received by typing it into the text box. Then, he/she chooses one of the ten options. **The option chosen by Player 2 determines the**

earnings of all players. Remember that if Player 2 chooses a zero-payment option, the final earnings will be €0 for him/her and 4€ for Player 1.

A1.2. Instructions for the 2-Sender game in Bitter Pill

You are participating in a study on economic decision-making. You have already earned **€5** for showing up on time. Please read these instructions carefully as they describe how you can earn **additional** money. You will be paid all your earnings in cash.

Please do not talk or communicate with other participants in any way. If you have questions, raise your hand and one of us will help you.

In the study, all participants are randomly assigned to groups of three. Within each group, the computer randomly assigns participants to the roles of **Player 1**, **Player 2**, and **Player 3**. You will be informed of your role on the computer screen.

Summary of the study

- There are ten options with payments for each player. Player 1 and Player 2 are informed of the payment each player receives in each option. On the other hand, Player 3 does not receive this information.
- **Player 1 chooses one message out of the two available messages** to be sent to Player 3. Each message states that a specific option is the option that gives the highest payment to Player 3.
- Which message will finally be delivered depends on a **private agreement between Players 1 and 2.** The agreement specifies an amount of money that Player 1 transfers to Player 2 for the delivery.
- **Player 2 delivers the message to Player 3 in person.**
- **Player 3 chooses an option** that determines the earnings of all players.

Specific instructions

There are ten options, each one labelled with a unique letter: A, B, C, D, E, F, G, H, I, or J. The computer will randomly assign one option to pay **€10 to Player 1, €10 to Player 2, and €10 to Player 3** and another option to pay **€17 to Player 1, €10 to Player 2, and €3 to Player 3**. The remaining eight options pay **€4 to Player 1, €4 to Player 2, and €0 to Player 3**.

How much each player earns in each option will be shown only to Player 1 and Player 2. The following table is an example of how payments could be assigned to the various options and how this information would be presented to Player 1 and Player 2.

Option	A	B	**C**	D	**E**	F	G	H	I	J
Player 1’s payment	4	4	10	4	17	4	4	4	4	4
Player 2’s payment	4	4	10	4	10	4	4	4	4	4
Player 3’s payment	0	0	10	0	3	0	0	0	0	0

Player 3 will not know which options provide positive earnings for him/her. The table below shows what Player 3 will see.

Option	A	B	C	D	E	F	G	H	I	J
Player 1’s payment	?	?	?	?	?	?	?	?	?	?
Player 2’s payment	?	?	?	?	?	?	?	?	?	?
Player 3’s payment	?	?	?	?	?	?	?	?	?	?

The only information that Player 3 receives regarding the payments of the various options is the message chosen by Player 1 and delivered to Player 3 by Player 2. After receiving the message, Player 3 chooses one of the ten options. The option chosen by Player 3 determines the earnings of all players in the group.

Player 1 chooses a message and reaches an agreement with Player 2

Player 1 chooses **one message** for Player 3. There are **two available messages**. Each message corresponds to one of the two options with positive earnings for all players.

Message I corresponds to the option that pays €10 to Player 3. The message reads “Option *<letter of option that pays €10 to Player 3>* will earn you 10 euros”.

Message II corresponds to the option that pays €3 to Player 3. The message reads “Option *<letter of option that pays €3 to Player 3>* will earn you 10 euros”.

Note that Player 1 cannot choose a message that corresponds to an option that pays €0 to Player 3. Therefore, when Player 3 receives a message, he/she will not know whether the option mentioned in the message pays him/her €10 or €3, but he/she can be certain that the option does not pay him/her €0.

Example

Suppose that the computer randomly assigns payments to options as shown in the table below.

Option	A	B	C	**D**	E	**F**	G	H	I	J
Player 1’s payment	4	4	4	17	4	10	4	4	4	4
Player 2’s payment	4	4	4	10	4	10	4	4	4	4
Player 3’s payment	0	0	0	3	0	10	0	0	0	0

In this case, Player 3 can receive one of the following two messages:

“Option F will earn you 10 euros”

“Option D will earn you 10 euros”

Player 1 cannot deliver the message to Player 3. Only Player 2 is able to deliver the message for him/her. If the option mentioned in the message coincides with the option subsequently chosen by Player 3, then **Player 1 transfers between €0 and €6.50 to Player 2 for delivery**. The screens below will be used to determine which message is delivered and how much is transferred. Each screen displays a list containing 14 rows, each row representing a possible transfer from Player 1 to Player 2. Player 1 and Player 2 must decide between Message I and Message II in each of the 14 rows. Players 1 and 2 make their 14 decisions **simultaneously**, which means that Player 2 will not know Player 1’s decisions while he/she is deciding, and vice-versa for Player 1. Specifically, in each row, Player 1 decides between:

- Choosing **Message I** and transferring **€0** to Player 2.
- Choosing **Message II** and transferring **the amount specified in that row** to Player 2.

Similarly, in each row, Player 2 decides between:

- Delivering **Message I** in exchange for a transfer from Player 1 of **€0**.
- Delivering **Message II** in exchange for a transfer from Player 1 equal to **the amount specified in that row**.

After both players have made their decisions, **one of the 14 rows will be randomly selected** by the computer to determine which message will be delivered to Player 3. All rows have the same chance of being selected; therefore, you should make your decision in each row seriously.

Decisions of Player 1

You are Player 1

Option	A	B	C	D	E	F	G	H	I	J
Player 1's payment	10	4	4	4	4	4	4	17	4	4
Player 2's payment	10	4	4	4	4	4	4	10	4	4
Player 3's payment	10	0	0	0	0	0	0	3	0	0

Please decide between Message I and Message II in **each row.**

Message I:
Option A will earn you 10 euros
Player 3 earns **€10** if he/she chooses Option A.

Message II:
Option H will earn you 10 euros
Player 3 earns **€3** if he/she chooses Option H.

Row	Transfer	Your payment	Earnings Player 2		Transfer	Your payment	Earnings Player 2
1	0	10	10	○ ○	0.0	17.0	10.0
2	0	10	10	○ ○	0.5	16.5	10.5
3	0	10	10	○ ○	1.0	16.0	11.0
4	0	10	10	○ ○	1.5	15.5	11.5
5	0	10	10	○ ○	2.0	15.0	12.0
6	0	10	10	○ ○	2.5	14.5	12.5
7	0	10	10	○ ○	3.0	14.0	13.0
8	0	10	10	○ ○	3.5	13.5	13.5
9	0	10	10	○ ○	4.0	13.0	14.0
10	0	10	10	○ ○	4.5	12.5	14.5
11	0	10	10	○ ○	5.0	12.0	15.0
12	0	10	10	○ ○	5.5	11.5	15.5
13	0	10	10	○ ○	6.0	11.0	16.0
14	0	10	10	○ ○	6.5	10.5	16.5

Decisions of Player 2

You are Player 2

Option	A	B	C	D	E	F	G	H	I	J
Player 1's payment	10	4	4	4	4	4	4	17	4	4
Player 2's payment	10	4	4	4	4	4	4	10	4	4
Player 3's payment	10	0	0	0	0	0	0	3	0	0

Please decide between Message I and Message II in **each row.**

Message I:
Option A will earn you 10 euros
Player 3 earns **€10** if he/she chooses Option A.

Message II:
Option H will earn you 10 euros
Player 3 earns **€3** if he/she chooses Option H.

Row	Transfer	Earnings Player 1	Your payment		Transfer	Earnings Player 1	Your payment
1	0	10	10	○ ○	0.0	17.0	10.0
2	0	10	10	○ ○	0.5	16.5	10.5
3	0	10	10	○ ○	1.0	16.0	11.0
4	0	10	10	○ ○	1.5	15.5	11.5
5	0	10	10	○ ○	2.0	15.0	12.0
6	0	10	10	○ ○	2.5	14.5	12.5
7	0	10	10	○ ○	3.0	14.0	13.0
8	0	10	10	○ ○	3.5	13.5	13.5
9	0	10	10	○ ○	4.0	13.0	14.0
10	0	10	10	○ ○	4.5	12.5	14.5
11	0	10	10	○ ○	5.0	12.0	15.0
12	0	10	10	○ ○	5.5	11.5	15.5
13	0	10	10	○ ○	6.0	11.0	16.0
14	0	10	10	○ ○	6.5	10.5	16.5

Player 2 will deliver the message determined by the choices in the selected row in the following way:

- In the selected row, if **Player 1 chooses Message I**, then regardless Player 2's choice, **Player 2 delivers Message I**. In this case, if Player 3 chooses the option corresponding to Message I, then Player 1, Player 2, and Player 3 all earn €10.
- In the selected row, if **Player 2 chooses Message I**, then regardless Player 1's choice, **Player 2 delivers Message I**. In this case, if Player 3 chooses the option corresponding to Message I, then Player 1, Player 2, and Player 3 all earn €10.
- In the selected row, if **both Player 1 and Player 2 choose Message II**, then **Player 2 delivers Message II**. In this case, if Player 3 chooses the option corresponding to Message II, then Player 1 earns €17 minus the transferred amount specified in that row, Player 2 earns €10 plus the transferred amount specified in that row, and Player 3 earns €3.

To summarize, Message II is delivered to Player 3 only when both Player 1 and Player 2 choose Message II in the selected row; otherwise Message I is delivered.

Player 3 will **not** be informed which row was selected by the computer.

Example

Suppose that Player 1 and Player 2 make the choices shown below.

Decisions of Player 1

You are Player 1

Option	A	B	C	D	E	F	G	H	I	J
Player 1's payment	10	4	4	4	4	4	4	17	4	4
Player 2's payment	10	4	4	4	4	4	4	10	4	4
Player 3's payment	10	0	0	0	0	0	0	3	0	0

Please decide between Message I and Message II in **each row.**

Message I: Option A will earn you 10 euros

Player 3 earns **€ 10** if he/she chooses Option A.

Message II: Option H will earn you 10 euros

Player 3 earns **€ 3** if he/she chooses Option H.

Row	Transfer	Your payment	Earnings Player 2		Transfer	Your payment	Earnings Player 2
1	0	10	10	○ ●	0.0	17.0	10.0
2	0	10	10	○ ●	0.5	16.5	10.5
3	0	10	10	○ ●	1.0	16.0	11.0
4	0	10	10	○ ●	1.5	15.5	11.5
5	0	10	10	○ ●	2.0	15.0	12.0
6	0	10	10	○ ●	2.5	14.5	12.5
7	0	10	10	○ ●	3.0	14.0	13.0
8	0	10	10	○ ●	3.5	13.5	13.5
9	0	10	10	○ ●	4.0	13.0	14.0
10	0	10	10	○ ●	4.5	12.5	14.5
11	0	10	10	● ○	5.0	12.0	15.0
12	0	10	10	● ○	5.5	11.5	15.5
13	0	10	10	● ○	6.0	11.0	16.0
14	0	10	10	● ○	6.5	10.5	16.5

Decisions of Player 2

You are Player 2

Option	A	B	C	D	E	F	G	H	I	J
Player 1's payment	10	4	4	4	4	4	4	17	4	4
Player 2's payment	10	4	4	4	4	4	4	10	4	4
Player 3's payment	10	0	0	0	0	0	0	3	0	0

Please decide between Message I and Message II in **each row.**

Message I: Option A will earn you 10 euros

Player 3 earns **€ 10** if he/she chooses Option A.

Message II: Option H will earn you 10 euros

Player 3 earns **€ 3** if he/she chooses Option H.

Row	Transfer	Earnings Player 1	Your payment		Transfer	Earnings Player 1	Your payment
1	0	10	10	● ○	0.0	17.0	10.0
2	0	10	10	● ○	0.5	16.5	10.5
3	0	10	10	● ○	1.0	16.0	11.0
4	0	10	10	● ○	1.5	15.5	11.5
5	0	10	10	● ○	2.0	15.0	12.0
6	0	10	10	○ ●	2.5	14.5	12.5
7	0	10	10	○ ●	3.0	14.0	13.0
8	0	10	10	○ ●	3.5	13.5	13.5
9	0	10	10	○ ●	4.0	13.0	14.0
10	0	10	10	○ ●	4.5	12.5	14.5
11	0	10	10	○ ●	5.0	12.0	15.0
12	0	10	10	○ ●	5.5	11.5	15.5
13	0	10	10	○ ●	6.0	11.0	16.0
14	0	10	10	○ ●	6.5	10.5	16.5

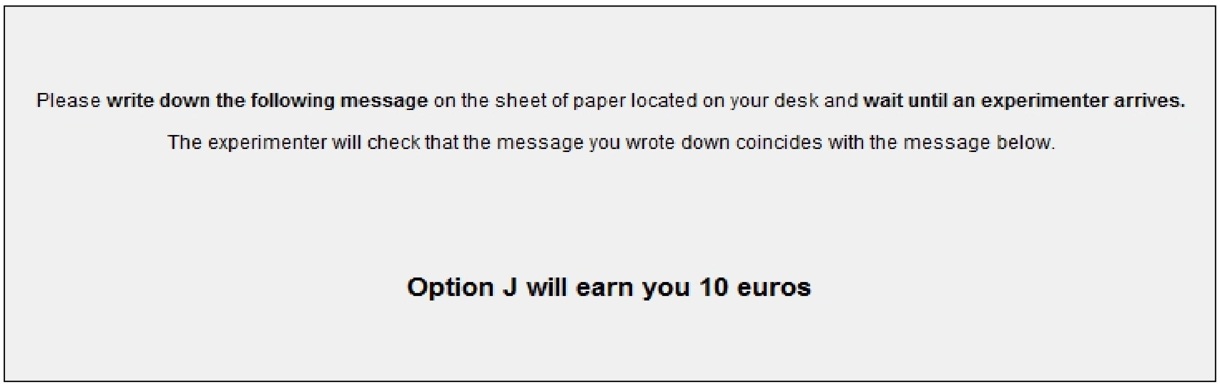
In this example, Player 1 is willing to transfer at maximum €4.5 to Player 2 for delivering Message II, while Player 2 demands at least €2.5 for delivering Message II. Given these choices, the following occurs if the computer randomly selects one of the rows below:

- **Row 4:** Since Player 1 chose Message II but Player 2 disagreed in favor of Message I, then Player 2 delivers Message I. Thereafter, if Player 3 chooses the option corresponding to Message I, then Player 1, Player 2, and Player 3 all earn €10.
- **Row 9:** Since Player 1 chose Message II and Player 2 agreed to Message II then Player 2 delivers Message II. Thereafter, if Player 3 chooses the option corresponding to Message II, then Player 1 earns €17 – €4 = €13, Player 2 earns €10 + €4 = €14, and Player 3 earns €3.

- **Row 12:** Since Player 1 chose Message I then Player 2 delivers Message I automatically. Thereafter, if Player 3 chooses the option corresponding to Message I, then Player 1, Player 2, and Player 3 all earn €10.

Player 2 delivers the message to Player 3 in person

Once the message is determined, Player 2 will see a screen like the one below.



To deliver the message, Player 2 will first **write down the message on the sheet of paper** located on his/her desk. Then, Player 2 will wait until an experimenter arrives. The experimenter will check whether the message written on the sheet of paper is identical to the message shown on the screen. Note that, like Player 3, the experimenter will not know to which payment the option in the message corresponds.

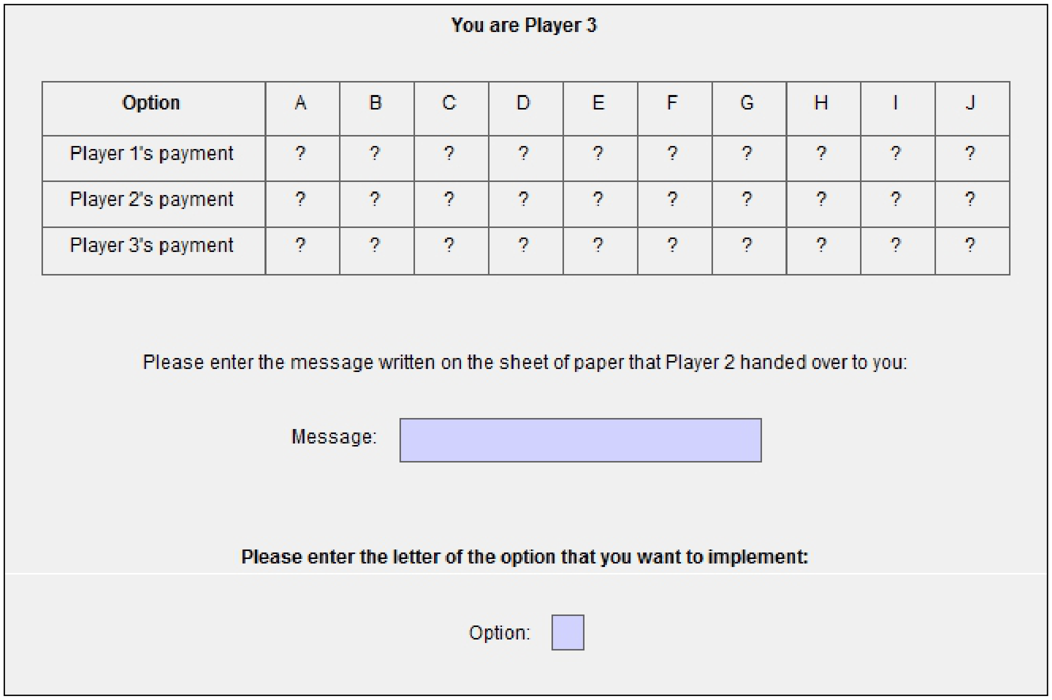
The experimenter will then walk with Player 2 to the desk of the Player 3 of his/her group. At this point, Player 2 will hand the paper with the message to Player 3 and then walk back to his/her desk.

Remember that **any kind of communication between the players is prohibited**, including gestures and facial expressions. In addition, Player 2 is not allowed to write down anything else other than the message on the sheet of paper. Any participant who does not comply with these rules will **not be paid** at the end of the study.

Player 3 chooses an option

Player 3 knows that there are two options with positive payments for him/her, but he/she does not know which two of the ten options contain these payments. **The only information that Player 3**

receives is the message delivered to him/her by Player 2. After receiving the message, Player 3 sees a screen like this:


You are Player 3

Option	A	B	C	D	E	F	G	H	I	J
Player 1's payment	?	?	?	?	?	?	?	?	?	?
Player 2's payment	?	?	?	?	?	?	?	?	?	?
Player 3's payment	?	?	?	?	?	?	?	?	?	?

Please enter the message written on the sheet of paper that Player 2 handed over to you:

Message:

Please enter the letter of the option that you want to implement:

Option:

On this screen, Player 3 first confirms the message he/she received by typing it into the text box. Then, he/she chooses one of the ten options. **The option chosen by Player 3 determines the earnings of all players.** Remember that if Player 3 chooses a zero-payment option, the final earnings will be €0 for him/her and 4€ for Player 1 and 2.

A1.3. Instructions for the Passive-Sender-B game in Deception

You are participating in a study on economic decision-making. You have already earned **€5** for showing up on time. Please read these instructions carefully as they describe how you can earn **additional** money. You will be paid all your earnings in cash.

Please do not talk or communicate with other participants in any way. If you have questions, raise your hand and one of us will help you.

In the study, all participants are randomly assigned to groups of three. Within each group, the computer randomly assigns participants to the roles of **Player 1**, **Player 2**, and **Player 3**. You will be informed of your role on the computer screen.

Summary of the study

- There are ten options with payments for each player. Player 1 and Player 2 are informed of the payment each player receives in each option. Player 3 does not receive this information.
- **Player 1 chooses one message out of the two available messages** to be sent to Player 3, each of them stating that a specific option gives the highest payment to Player 3.
- By choosing the message, Player 1 specifies an amount of money that he/she transfers to Player 2 for the delivery.
- **Player 2 delivers the message to Player 3 in person.**
- **Player 3 chooses an option** that determines the earnings of all players.

Specific instructions

There are ten options, each labelled with a unique letter: A, B, C, D, E, F, G, H, I, or J. The computer will randomly assign one option to pay **€10 to Player 1, €10 to Player 2, and €10 to Player 3**. It will also randomly assign another option to pay **€17 to Player 1, €10 to Player 2, and €3 to Player 3**. The remaining eight options pay **€4 to Player 1, €4 to Player 2, and €0 to Player 3**.

Importantly, **how much each player earns in each option will be shown only to Player 1 and Player 2**. The following table is an example of how payments could be assigned to the various options and how this information would be presented to Player 1 and Player 2.

Option	A	B	**C**	D	**E**	F	G	H	I	J
Player 1's payment	4	4	10	4	17	4	4	4	4	4
Player 2's payment	4	4	10	4	10	4	4	4	4	4
Player 3's payment	0	0	10	0	3	0	0	0	0	0

By contrast, **Player 3 will not know which options provide positive earnings for him/her.** The table below shows what Player 3 will see instead.

Option	A	B	C	D	E	F	G	H	I	J
Player 1's payment	?	?	?	?	?	?	?	?	?	?
Player 2's payment	?	?	?	?	?	?	?	?	?	?
Player 3's payment	?	?	?	?	?	?	?	?	?	?

The only information that Player 3 receives regarding the payments of the various options is a message chosen by Player 1 and delivered to him/her by Player 2. After receiving the message, Player 3 implements one of the ten options. The option implemented by Player 3 determines the earnings of all players.

Player 1 chooses a message

Player 1 chooses **one message** for Player 3. There are **two available messages**. Each message corresponds to one of the two options with positive earnings for all players.

Message I corresponds to the option that pays €10 to Player 3. It reads "Option *[letter of option that pays €10 to Player 3]* will earn you 10 euros".

Message II corresponds to the option that pays €3 to Player 3. It reads "Option *[letter of option that pays €3 to Player 3]* will earn you 10 euros".

Note that Player 1 cannot choose a message that corresponds to an option that pays €0 to Player 3. Therefore, when Player 3 receives a message, he/she will not know whether the option mentioned in the message pays him/her €10 or €3, but he/she can be certain that the option does not pay him/her €0.

Example

Suppose that the computer randomly assigns payments to options as shown in the table below.

Option	A	B	C	**D**	E	**F**	G	H	I	J
Player 1's payment	4	4	4	17	4	10	4	4	4	4
Player 2's payment	4	4	4	10	4	10	4	4	4	4
Player 3's payment	0	0	0	3	0	10	0	0	0	0

In this case, Player 3 can receive one of the following two messages:

"Option D will earn you 10 euros"

"Option F will earn you 10 euros"

Player 1 cannot deliver the message directly to Player 3. Instead, Player 2 delivers the message for him/her. If the delivered message is subsequently implemented by Player 3, then **Player 1 transfers**

between €0 and €6.50 to Player 2 for delivery. The screens below will be used to determine which message is delivered and how much is transferred. Each screen displays a list containing 14 rows, each row representing a possible transfer from Player 1 to Player 2. Player 1 must decide between Message I and Message II in each of the 14 rows.

Decisions of Player 1

You are Player 1

Option	A	B	C	D	E	F	G	H	I	J
Player 1's payment	10	4	4	4	4	4	4	17	4	4
Player 2's payment	10	4	4	4	4	4	4	10	4	4
Player 3's payment	10	0	0	0	0	0	0	3	0	0

Please decide between Message I and Message II in **each row.**

Message I:
Option A will earn you 10 euros

Player 3 earns **€ 10** if Option A is implemented.

Message II:
Option H will earn you 10 euros

Player 3 earns **€ 3** if Option H is implemented.

Row	Transfer	Your payment	Earnings Player 2		Transfer	Your payment	Earnings Player 2
1	0	10	10	○ ○	0.0	17.0	10.0
2	0	10	10	○ ○	0.5	16.5	10.5
3	0	10	10	○ ○	1.0	16.0	11.0
4	0	10	10	○ ○	1.5	15.5	11.5
5	0	10	10	○ ○	2.0	15.0	12.0
6	0	10	10	○ ○	2.5	14.5	12.5
7	0	10	10	○ ○	3.0	14.0	13.0
8	0	10	10	○ ○	3.5	13.5	13.5
9	0	10	10	○ ○	4.0	13.0	14.0
10	0	10	10	○ ○	4.5	12.5	14.5
11	0	10	10	○ ○	5.0	12.0	15.0
12	0	10	10	○ ○	5.5	11.5	15.5
13	0	10	10	○ ○	6.0	11.0	16.0
14	0	10	10	○ ○	6.5	10.5	16.5

Similarly, in each row, Player 2 decides between:

- Delivering **Message I** and transferring **€0** to Player 2.
- Delivering **Message II** and transferring **the amount specified in that row** to Player 2.

After Player 1 has made his/her decisions, **one of the 14 rows will be randomly selected** by the computer to determine which message will be delivered to Player 3. All rows have the same chance of being selected; therefore, you should make your decision in each row seriously. Player 2 will deliver the message determined by the choices in the selected row in the following way:

- In the selected row, if **Player 1 chooses Message I**, then **Player 2 delivers Message I**. In this case, if Player 3 implements the option in Message I, then Player 1, Player 2, and Player 3 all earn €10.

- In the selected row, if **Player 1 chooses Message II**, then **Player 2 delivers Message II**. In this case, if Player 3 implements the option in Message II, then Player 1 earns €17 minus the transferred amount specified in that row, Player 2 earns €10 plus the transferred amount specified in that row, and Player 3 earns €3.

To summarize, Message II is delivered to Player 3 only when Player 1 has chosen Message II in the selected row; otherwise Message 1 is delivered.

Player 3 will **not** be informed which row was selected by the computer.

Examples

Suppose that Player 1 makes the choices shown below.

Decisions of Player 1

You are Player 1

Option	A	B	C	D	E	F	G	H	I	J
Player 1's payment	10	4	4	4	4	4	4	17	4	4
Player 2's payment	10	4	4	4	4	4	4	10	4	4
Player 3's payment	10	0	0	0	0	0	0	3	0	0

Please decide between Message I and Message II in **each row.**

Message I:
Option A will earn you 10 euros

Player 3 earns **€ 10** if Option A is implemented.

Message II:
Option H will earn you 10 euros

Player 3 earns **€ 3** if Option H is implemented.

Row	Transfer	Your payment	Earnings Player 2		Transfer	Your payment	Earnings Player 2
1	0	10	10	○ ◉	0.0	17.0	10.0
2	0	10	10	○ ◉	0.5	16.5	10.5
3	0	10	10	○ ◉	1.0	16.0	11.0
4	0	10	10	○ ◉	1.5	15.5	11.5
5	0	10	10	○ ◉	2.0	15.0	12.0
6	0	10	10	○ ◉	2.5	14.5	12.5
7	0	10	10	○ ◉	3.0	14.0	13.0
8	0	10	10	○ ◉	3.5	13.5	13.5
9	0	10	10	○ ◉	4.0	13.0	14.0
10	0	10	10	○ ◉	4.5	12.5	14.5
11	0	10	10	◉ ○	5.0	12.0	15.0
12	0	10	10	◉ ○	5.5	11.5	15.5
13	0	10	10	◉ ○	6.0	11.0	16.0
14	0	10	10	◉ ○	6.5	10.5	16.5

In this example, Player 1 is willing to transfer at maximum €4.5 to Player 2 for delivering Message II. Given these choices, the following occurs if the computer randomly selects:

- **Row 9:** Since Player 1 chose Message II then Player 2 delivers Message II. Thereafter, if Player 3 implements the option associated to Message II, then Player 1 earns €17 – €4 = €13, Player 2 earns €10 + €4 = €14, and Player 3 earns €3.

- **Row 12:** Since Player 1 chose Message I then Player 2 delivers Message I. Thereafter, if Player 3 implements the option associated to Message I, then Player 1, Player 2, and Player 3 earn €10.

Suppose that Player 1 makes the choices shown below.

Decisions of Player 1

You are Player 1

Option	A	B	C	D	E	F	G	H	I	J
Player 1's payment	10	4	4	4	4	4	4	17	4	4
Player 2's payment	10	4	4	4	4	4	4	10	4	4
Player 3's payment	10	0	0	0	0	0	0	3	0	0

Please decide between Message I and Message II in **each row.**

Message I:
Option A will earn you 10 euros

Player 3 earns **€ 10** if Option A is implemented.

Message II:
Option H will earn you 10 euros

Player 3 earns **€ 3** if Option H is implemented.

Row	Transfer	Your payment	Earnings Player 2		Transfer	Your payment	Earnings Player 2
1	0	10	10	○ ●	0.0	17.0	10.0
2	0	10	10	○ ●	0.5	16.5	10.5
3	0	10	10	○ ●	1.0	16.0	11.0
4	0	10	10	○ ●	1.5	15.5	11.5
5	0	10	10	○ ●	2.0	15.0	12.0
6	0	10	10	● ○	2.5	14.5	12.5
7	0	10	10	● ○	3.0	14.0	13.0
8	0	10	10	● ○	3.5	13.5	13.5
9	0	10	10	● ○	4.0	13.0	14.0
10	0	10	10	● ○	4.5	12.5	14.5
11	0	10	10	● ○	5.0	12.0	15.0
12	0	10	10	● ○	5.5	11.5	15.5
13	0	10	10	● ○	6.0	11.0	16.0
14	0	10	10	● ○	6.5	10.5	16.5

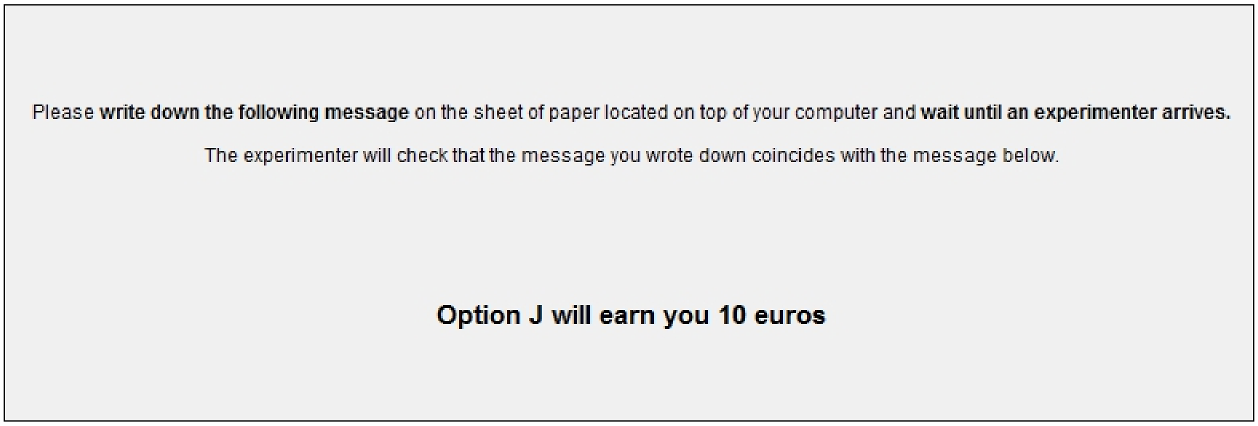
In this example, Player 1 is willing to transfer at maximum €2 to Player 2 for delivering Message II. Given these choices, the following occurs if the computer randomly selects:

- **Row 4:** Since Player 1 chose Message II then Player 2 delivers Message II. Thereafter, if Player 3 implements the option associated to Message II, then Player 1 earns 17€ - 1.5€ = 15.5€, Player 2 earns €10 + €1.5 = €11.5, and Player 3 earns €3.
- **Row 6:** Since Player 1 chose Message I then Player 2 delivers Message I. Thereafter, if Player 3 implements the option associated to Message I, then Player 1, Player 2, and Player 3 earn €10.

Player 2 delivers the message to Player 3 in person

Once the message is determined, Player 2 will see a screen like the one below.

Decisions of Player 1



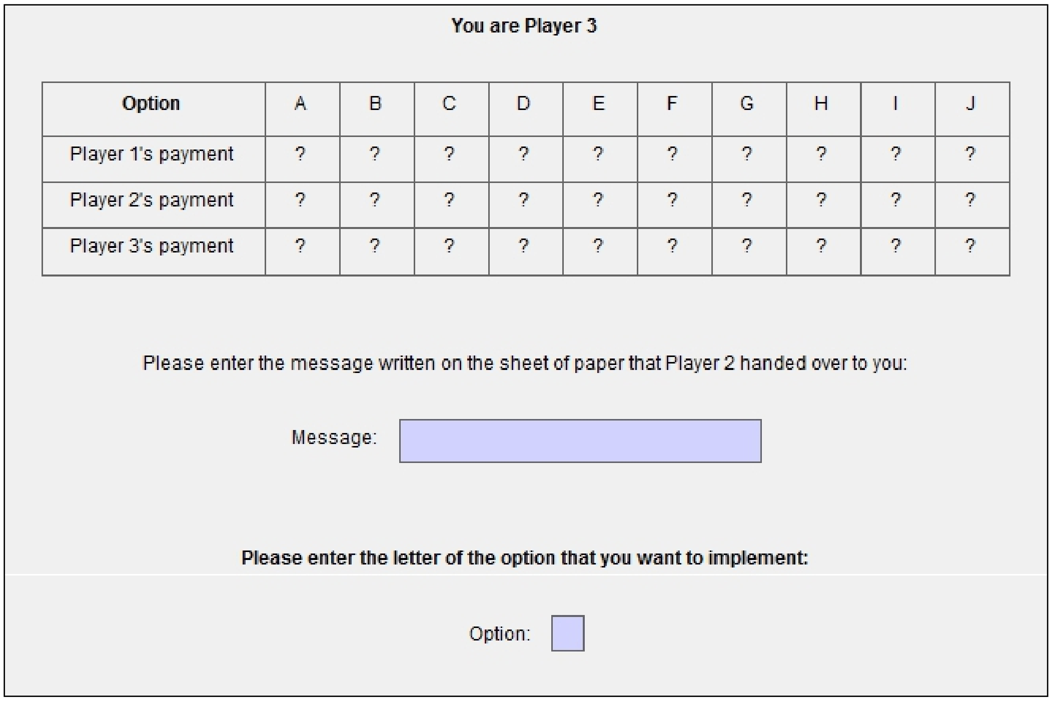
To deliver the message, Player 2 will first **write down the message on the sheet of paper** located on his/her desk. Then, Player 2 will wait until an experimenter arrives. The experimenter will check whether the written message is identical to the message shown on the screen. Note that, like Player 3, the experimenter will not know to which payment the option corresponds.

The experimenter will then walk with Player 2 to the desk where the Player 3 with whom he/she is matched with is sitting. At this point, **Player 2 will hand the paper with the message to Player 3** and then walk back to his/her desk.

Remember that **any kind of communication between the players is prohibited**, including gestures and facial expressions. In addition, Player 2 is not allowed to write down anything else other than the message on the sheet of paper. Any participant who does not comply with these rules will **not be paid** at the end of the study.

Player 3 implements an option

Player 3 knows that there are two options with positive payments for him/her, but he/she does not know which two of the ten options contain these payments. **The only information that Player 3 receives is the message delivered to him/her by Player 2**. After receiving the message, Player 3 sees a screen like this:


You are Player 3

Option	A	B	C	D	E	F	G	H	I	J
Player 1's payment	?	?	?	?	?	?	?	?	?	?
Player 2's payment	?	?	?	?	?	?	?	?	?	?
Player 3's payment	?	?	?	?	?	?	?	?	?	?

Please enter the message written on the sheet of paper that Player 2 handed over to you:

Message:

Please enter the letter of the option that you want to implement:

Option:

In this screen, Player 3 first confirms the message he/she received by typing it in. Then, he/she chooses one of the ten options to implement. **The option implemented by Player 3 determines the earnings of all players.** Remember that if Player 3 implements a zero-payment option, the final earnings will be €0 for him/her and 4€ for Player 1 and 2.

A2. Complementary statistical analyses

A2.1. Treatment comparisons

Table A1 shows the regressions used to evaluate whether the treatment differences in the senders' antisocial premiums are statistically significant. The coefficients are estimated using interval regressions. Senders who choose Message II over Message I when they earn at least €*x* are classified as having an antisocial premium in the interval [€*x* – 0.5, €*x*]. Senders who always choose Message I are classified as having an antisocial premium in the interval [€7.50, ∞) if they played in the *1-Sender* game or in the *2-Sender* game as Sender A, and [€7.00, ∞) if they played the *2-Sender* game as Sender B. Senders who always choose Message II are classified as having an antisocial premium in the interval (∞, €0.50] if they played in the *1-Sender* game or in the *2-Sender* game as Sender A, and (∞, €0.00] if they played the *2-Sender* game as Sender B. All regressions are estimated using robust

Table A1. Treatment differences in antisocial premiums

Note: Interval regressions of the senders' antisocial premium. Robust standard errors in parentheses. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	I	II	III
2-Sender	–1.58***		
	(0.57)		
2-Sender × Bitter pill		–1.57***	-1.57**
		(0.73)	(0.74)
2-Sender × Deception		–1.57***	-1.59**
		(0.78)	(0.79)
Passive-Sender-B			0.56
			(0.95)
Deception		–1.63*	–1.63*
		(0.91)	(0.93)
Constant	3.99***	4.82***	4.82***
	(0.49)	(0.60)	(0.61)
Observations	110	110	131
χ^2	7.73	24.32	25.92

standard errors. The regressions in columns I and II use data from to the *1-Sender* and the *2-Sender* games. Column III further includes the data from the *Passive-Sender-B* game.

Table A2 shows the means and standard deviations of selected variables used in the analyses. The first three columns contain the respective values for the *1-Sender* and *2-Sender* games with pooled data from the Bitter pill and Deception contexts as well as the values for the *Passive-Sender-B* game, which is only applied to the *Deception* context. Columns three to six report these values in the *1-Sender* and *2-Sender* games separately for each context. Note that senders' normative views of sending Message I are elicited on a five-point Likert scale ranging from very unacceptable (1) to very acceptable (5). Moreover, senders' additional earnings if the receiver follows Message II refers to the surplus that senders gain in the selected row compared to the equal payoff distribution, assuming that Message II is sent and that the receiver implements the option mentioned in this message. Specifically, this amount equals seven minus *x* for the sender in the *1-Sender* game as well as Sender A in the *2-Sender* and *Passive-Sender-B* games and to *x* for Sender B in the *2-Sender* and *Passive-Sender-B* games.

Table A2. Means and standard deviations of selected variables by game and treatment

	Overall			Bitter pill		Deception	
	1-Sender	*2-Sender*	*Passive-Sender-B*	*1-Sender*	*2-Sender*	*1-Sender*	*2-Sender*
Fraction of Message IIs sent	0.44	0.29	0.46	0.38	0.11	0.50	0.47
	(0.50)	(0.46)	(0.5)	(0.49)	(0.31)	(0.51)	(0.50)
Senders' normative views of sending Message I	4.51	4.31	4.37	4.58	4.2	4.45	4.42
	(0.85)	(1.04)	(1.00)	(0.69)	(1.21)	(1.00)	(0.84)
Senders' additional earnings if the receiver follows Message II	3.31	3.49	3.57	3.66	3.51	2.98	3.47
	(2.07)	(1.92)	(2.05)	(2.17)	(1.85)	(1.97)	(2.01)
Fraction of receivers who followed Message II	0.89	0.97	1.00	0.77	0.95	1.00	1.00
	(0.32)	(0.16)	(0.00)	(0.43)	(0.23)	(0.00)	(0.00)
Senders' expected fraction of receivers following Message II	0.84	0.83	0.85	0.84	0.81	0.85	0.85
	(0.21)	(0.23)	(0.25)	(0.24)	(0.26)	(0.2)	(0.21)
Fraction of women	0.49	0.41	0.46	0.46	0.36	0.53	0.45
	(0.50)	(0.49)	(0.50)	(0.51)	(0.49)	(0.51)	(0.50)
Age	23.63	22.74	21.85	22.87	22.8	24.38	22.67
	(5.61)	(2.46)	(4.52)	(3.99)	(2.44)	(6.80)	(2.51)

Table A3 shows the regressions used to evaluate whether the treatment differences in the subjects' normative views are statistically significant. All normative views range from very unacceptable (1) to very acceptable (5). Therefore, we estimate all coefficients using ordered probit regressions with robust standard errors. Moreover, the standard errors of senders in the *2-Sender* game are clustered on their matched pairs. In columns I to III, the dependent variable is the senders' normative views concerning the acceptability of sending Message II. In columns IV to VI, the dependent variable is the senders' belief of the receivers' normative views concerning the acceptability of sending Message II. Lastly, in columns VII to IX, the dependent variable is the receivers' normative views concerning the acceptability of sending Message II. All regressions include date from the *1-Sender* and *2-Sender* games. Regressions in columns III, VI, and IX further include the data from the *Passive-Sender-B* game.

Table A4 provides robustness checks for the results from the *1-Sender* and the *2-Sender* games observed in Table A3. Ordered probit regressions are used with robust standard errors that are clustered on their matched pairs in the *2-Sender* game. In columns I and II, the dependent variable is the senders' normative views concerning the acceptability of sending Message II. Unlike in Table A3,

Table A3. Treatment differences in normative views

Note: Ordered probit regressions of the subjects' normative views. Robust standard errors clustered on matched pairs (for senders in the *2-Sender* game) in parentheses. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	Senders' normative views			Senders' expectations about receivers' normative views			Receivers' normative views		
	I	II	III	IV	V	VI	VII	VIII	IX
2-Sender game	0.62***			0.65***			0.39		
	(0.23)			(0.24)			(0.24)		
2-Sender × Bitter pill		0.70**	0.66**		0.54	0.55		0.30	0.27
		(0.30)	(0.29)		(0.35)	(0.36)		(0.34)	(0.34)
2-Sender × Deception		0.56*	0.53*		0.76**	0.77**		0.53	0.54
		(0.34)	(0.32)		(0.33)	(0.33)		(0.34)	(0.34)
Deception		0.44	0.42		-0.25	-0.25		0.50*	0.49
		(0.36)	(0.34)		(0.37)	(0.38)		(0.30)	(0.30)
Sender A in			-0.01			0.20			
Passive-Sender-B			(0.37)			(0.34)			
Passive Sender B in			-0.31			0.38			
Passive-Sender-B			(0.34)			(0.36)			
Receiver in									-0.12
Passive-Sender-B									(0.32)
Observations	110	110	153	110	110	153	79	79	101
Clusters	77	77	99	77	77	99	79	79	101
χ^2	6.94	9.08	12.16	7.23	8.17	11.16	2.67	11.76	11.40

we also include the following control variables: a dummy variable that equals one if the other sender in the *2-Sender* game chose Message II, a dummy variable that equals one if the message sent to the receiver was Message II, and the sender's earnings if the receiver follows the message. Note that the control variables are neither jointly significant in column I (p = 0.968) nor column II (p = 0.844).

In columns III and IV, the dependent variable is the senders' belief of the receivers' normative views concerning the acceptability of sending Message II. These regressions also include the abovementioned control variables. Once again, note that the control variables are neither jointly significant in column III (p = 0.771) nor column IV (p = 0.837). Unlike in Table A4, in columns V and VI, the dependent variable is the senders' normative views concerning the acceptability of sending Message I.

Table A4. Treatment differences in normative views, robustness checks

Note: Ordered probit regressions of the subjects' normative views. Robust standard errors clustered on matched pairs (for senders in the *2-Sender* game) in parentheses. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	I	II	III	IV	V	VI
2-Sender	0.63**		0.58*		–0.26	
	(0.27)		(0.32)		(0.24)	
2-Sender × Bitter pill		0.75**		0.48		–0.40
		(0.33)		(0.40)		(0.33)
2-Sender × Deception		0.59		0.68*		–0.11
		(0.37)		(0.39)		(0.35)
Deception		0.50		–0.20		–0.10
		(0.37)		(0.38)		(0.39)
Other sender chose Message II	–0.01	–0.05	0.11	0.10		
	(0.28)	(0.29)	(0.30)	(0.30)		
Message II was sent	0.29	0.40	0.24	0.19		
	(0.59)	(0.62)	(0.58)	(0.60)		
Earnings if message is followed	0.04	0.08	0.06	0.06		
	(0.09)	(0.09)	(0.08)	(0.09)		
Observations	110	110	110	110	110	110
Clusters	77	77	77	77	77	77
χ^2	8.18	12.16	9.75	10.69	1.12	1.68

Table A5 shows the regressions used to evaluate whether the treatment differences in the subjects' beliefs about the probability that a receiver receives Message II (simply referred to as the senders' "second-order beliefs") are statistically significant. All beliefs are censored at 0% and 100%. Therefore, we estimate all coefficients using Tobit regressions with robust standard errors. Moreover, the standard errors of senders in the *2-Sender* game are clustered on their matched pairs. In columns I to III, the dependent variable is the senders' belief of the receivers' expected probability of receiving Message II. In columns IV and VI, the dependent variable is the receivers' expected probability of receiving Message II. All regressions include data from to the *1-Sender* and the *2-Sender* games. Regressions in columns III and VI further include the data from the *Passive-Sender-B* game.

Table A5. Treatment differences in senders' second-order beliefs and the receivers' own beliefs regarding the probability of receiving Message II

Note: Tobit regressions of the senders' belief of the receivers' expected probability of receiving Message II and the receivers' expected probability of receiving Message II. Robust standard errors clustered on matched pairs (for the *2-Sender* and *Passive-Sender-B* games) in parentheses. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	I	II	III	IV	V	VI
2-Sender	–0.03			–0.07		
	(0.08)			(0.07)		
2-Sender × Bitter pill		–0.05	-0.05		–0.27***	-0.27***
		(0.10)	(0.10)		(0.10)	(0.10)
2-Sender × Deception		0.00	0.00		0.13	0.13
		(0.12)	(0.12)		(0.09)	(0.09)
Deception		0.11	0.11		–0.03	-0.03
		(0.13)	(0.13)		(0.10)	(0.10)
Sender A in *Passive-Sender-B*			0.00			
			(0.13)			
Sender B in *Passive-Sender-B*			0.04			
			(0.12)			
Receiver in *Passive-Sender-B*						0.01
						(0.11)
Constant	0.59***	0.54***	0.54***	0.56***	0.58***	0.58***
	(0.07)	(0.09)	(0.09)	(0.05)	(0.07)	(0.07)
Observations	110	110	153	79	79	101
Clusters	77	77	99	79	79	101
F-statistic	0.11	1.51	1.73	0.96	6.31	4.73

A2.2. Antisocial premium regressions

Table A6 shows the regressions used to evaluate the effect of senders' normative views and second-order beliefs about the receivers' expected probability of receiving Message II on the antisocial premium, which is the dependent variable in all six regressions. We use interval regressions to account for the fact that when we observe a sender who switches from Message I to Message II when the latter pays more than €*x*, her antisocial premium lies in the interval [€*x* – 0.50, €*x*]. All regressions are estimated using standard errors clustered on matched pairs.

Table A6. Determinants of the antisocial premium

Note: Interval regressions of the senders' antisocial premium. Robust standard errors in parentheses. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	I	II	III	IV	V	VI	VII
Normative views	−0.39	−0.98*	−1.03*	−0.74	−0.20	−1.06***	−1.00**
	(0.33)	(0.50)	(0.55)	(0.51)	(0.25)	(0.39)	(0.39)
Second-order belief	−6.32***	−9.15***	−9.39***	−8.60***	−3.01***	−7.25***	−7.50***
	(1.48)	(2.68)	(3.02)	(2.67)	(0.91)	(1.80)	(1.83)
Normative views × Second-order belief		1.13	1.20	1.11		1.44***	1.31**
		(0.93)	(1.01)	(0.74)		(0.53)	(0.53)
Deception	−1.00	−0.94	−1.04	−1.06	−1.08**	−0.89**	−0.68*
	(0.76)	(0.76)	(0.77)	(0.79)	(0.44)	(0.39)	(0.36)
Expected probability of receiver following message			−0.15	−0.48			−0.11
			(0.32)	(0.31)			(0.20)
Female			0.36	0.12			0.06
			(0.38)	(0.36)			(0.23)
Age			0.38	1.05			0.29
			(0.67)	(2.44)			(0.31)
Age^2			−0.12	−1.09			−0.18
			(0.18)	(2.40)			(0.11)
Sender A in *Passive-Sender-B*				0.50			−1.55***
				(0.87)			(0.51)
Sender B in *2-Sender*			10.76***				10.29***
			(1.77)				(1.62)
Constant	9.02***	10.48***	−1.03*	9.66***	5.26***	7.66***	−1.00**
	(1.08)	(1.47)	(0.55)	(1.60)	(0.97)	(1.38)	(0.39)
Observations	39	39	39	60	71	71	71
χ^2	25.49	30.42	40.57	39.95	23.61	34.20	64.74

The regressions in columns I to III use data from the *1–Sender* game, the regression in column IV also includes data from the *Passive-Sender-B* game, and the regressions in columns V to VII use data from the *2-Sender* game. Regressions I and V include senders' normative views, second-order beliefs and an indicator variable for the context in which the game is played. In regressions II and VI we add the interaction of senders' normative views and second-order belief. In regressions III, IV, and VII we further control for senders' expected probability that the receiver follows the message, gender, and age. In regression IV, we include an indicator variable for being the Sender in the *Passive-Sender-B*

Table A7. Determinants of the antisocial premium, alternative specifications

Note: Interval regressions of the senders' antisocial premium. Robust standard errors in parentheses. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	I	II	III	IV	V	VI	VII
Second-order normative views	−0.10	0.95	0.91	−0.02	−0.05	−0.95***	−0.88***
	(0.28)	(0.88)	(1.04)	(0.86)	(0.16)	(0.28)	(0.31)
Second-order belief	−6.36***	−3.40	−3.26	−4.86*	−3.11***	−7.29***	−7.55***
	(1.59)	(3.30)	(2.88)	(2.78)	(0.95)	(1.37)	(1.54)
Second-order normative views × Second-order belief		−2.24	−2.34	−0.45		1.68***	1.53***
		(1.84)	(1.98)	(1.45)		(0.39)	(0.43)
Deception	−1.16	−1.23*	−1.39*	−1.29*	−1.15**	−1.14***	−0.93***
	(0.76)	(0.74)	(0.74)	(0.76)	(0.45)	(0.42)	(0.36)
Expected probability of receiver following message			−0.22	−0.51			−0.07
			(0.45)	(0.34)			(0.18)
Female			0.38	0.12			0.13
			(0.37)	(0.34)			(0.19)
Age			0.27	−0.13			0.24
			(0.74)	(0.57)			(0.25)
Age^2			−0.07	0.02			−0.21*
			(0.22)	(0.18)			(0.11)
Sender A in *Passive-Sender-B*				0.81			
				(0.85)			
Sender B in *2-Sender*							−1.46***
							(0.46)
Constant	8.37***	6.94***	7.14***	7.80***	4.87***	7.23***	9.72***
	(1.08)	(1.73)	(1.77)	(1.67)	(0.82)	(1.03)	(1.39)
Observations	39	39	39	60	71	71	71
χ^2	23.42	37.59	26.95	41.33	27.15	48.61	63.72

game, and in regression VII we include an indicator variable for being Sender B in the *2-Sender* game. Note that we decided to pool the data from the *Passive-Sender-B* game with the *1-Sender* game (as opposed to the *2-Sender* game) because these contexts display similar behavior.

Table A7 reports the coefficients of the same regressions used to evaluate the effect of senders' normative views and second-order beliefs on the antisocial premium as in Table A7 with two exceptions. First, in all seven regressions, we use the senders' second-order normative views instead of their own normative views, that is, the senders' estimation of how the receiver rated the

Table A8. Means and standard deviations of the senders' self-reported emotions

	1-Sender		*2-Sender*		*Passive-Sender-B*	
	Message I	Message II	Message I	Message II	Message I	Message II
Guilt	1.09	4.23	1.62	2.68	1.13	3.95
	(0.29)	(2.42)	(1.23)	(1.49)	(0.63)	(2.11)
Shame	1.27	3.85	1.46	2.16	1.22	2.3
	(0.77)	(2.76)	(0.93)	(1.80)	(0.52)	(1.72)
Anger	1.14	1.85	2.22	1.32	1.39	1.5
	(0.35)	(1.57)	(1.73)	(0.95)	(1.03)	(0.95)
Happiness	5.82	5.69	4.62	6.11	5.57	5.1
	(1.26)	(1.55)	(1.69)	(0.99)	(1.12)	(1.37)
Gratitude	5.77	5.38	4.60	5.79	5.3	3.9
	(1.48)	(1.71)	(1.80)	(1.08)	(1.89)	(1.83)

acceptability of sending Message II on a 5-point Likert scale ranging from "1 - very unacceptable" to "5 - very acceptable". Second, in regressions II, III, IV, VI and VII, we include the interaction of senders' second-order beliefs with their second-order normative views instead of with their own normative views.

A2.3. Emotions regressions

Table A8 provides the means and standard deviations of the senders' self-reported emotions dependent on whether Message I or Message II was sent to the receiver. Senders' emotions refer to the moment they learned the outcomes of all players and were elicited on a seven-point Likert scale ranging from 1 to 7 after the game was played.

Table A9 shows the regressions used to evaluate potential determinants of the level of guilt experienced by senders in the moment they learned the outcomes of all players. The guilt level was elicited on a seven-point Likert scale ranging from 1 to 7 after the game was played. In all models, we use linear regressions with robust standard errors clustered on matched pairs and the senders' experienced guilt as the dependent variable. Regressions I to III use data from the *1-Sender* game. In regression IV we pooled the data from the *Passive-Sender-B* game with the *1-Sender* game (as opposed to the *2-Sender* game) since these contexts display similar behavior. In regressions V to VII use data from the 2-*Sender* game. 'Delivered Message I' and 'Delivered Message II' are dummy variables indicating the message that was delivered to and followed by the receiver; 'Normative views' are the

Table A9. Determinants of experienced guilt

Note: Linear regressions of the senders' experienced guilt after observing the outcome of the game. Robust standard errors clustered on matched pairs (for senders in the *2-Sender* game) in parentheses. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	***1-Sender* game**				***2-Sender* game**		
	I	II	III	IV	V	VI	VII
Delivered Message II	3.14***	3.38***	6.30***	5.00***	1.06**	0.89**	2.76*
	(0.67)	(0.75)	(1.86)	(2.24)	(0.29)	(0.38)	(1.50)
Delivered Message I × Normative views			0.00	-0.2			0.00
			(0.15)	(0.14)			(0.19)
Delivered Message II × Normative views			–0.91***	-0.49***			–0.38*
			(0.29)	(0.38)			(0.22)
Delivered Message I × Second-order belief			0.18	-0.42			1.13
			(0.33)	(0.44)			(0.90)
Delivered Message II × Second-order belief			–0.65	-1.66			–0.25
			(2.79)	(2.52)			(1.39)
Deception		–0.04	–0.19	0.42		0.13	0.26
		(0.37)	(0.37)	(0.46)		(0.40)	(0.42)
Antisocial premium		–0.03	–0.06	-0.03		–0.05	0.05
		(0.12)	(0.11)	(0.13)		(0.06)	(0.12)

continued in the next page

senders' normative views of sending Message II; 'Second-order belief' is the senders' belief of the receivers' expected probability of receiving Message II. In regressions III, IV and VII, we add the interaction effects of normative views and, respectively, second-order beliefs with the message that was sent. Regressions II-IV and VI-VII include further controls for Message II being dishonest ('Deception'), the antisocial premium, the interaction of the message sent with the senders' earnings from Message II, gender and age. In regression IV we also include an indicator variable for being the Sender in the *Passive-Sender-B* game. Regressions VI and VII include an indicator variable for being Sender B in the *2-Sender* game.

In Table A10, we report the regressions used to evaluate potential determinants of the senders' experienced anger and shame. As in the case of guilt, these negative emotions refer to the moment senders learned the outcomes of all players and were elicited on a seven-point Likert scale ranging from 1 to 7 after the game was played. Like in Table A10, we use linear regressions with robust standard errors clustered on matched pairs in all models. In regressions I and II (*1-Sender*) as well as

Table A9. Determinants of experienced guilt (continued)

		***1-Sender* game**				***2-Sender* game**	
	I	II	III	IV	V	VI	VII
Delivered Message I × Sender's earnings if Message II is applied		–0.22	–0.05	-0.26		–0.27	–0.24
		(0.17)	(0.16)	(0.20)		(0.17)	(0.16)
Delivered Message II × Sender's earnings if Message II is applied		–1.48***	–1.37***	-0.54***		0.67	0.80
		(0.53)	(0.52)	(0.51)		(0.67)	(0.65)
Female		–0.25	–0.07	-0.06		0.06	0.12
		(0.22)	(0.21)	(0.18)		(0.14)	(0.14)
Age		0.69	0.15	0.44		0.03	0.06
		(0.46)	(0.34)	(0.44)		(0.19)	(0.24)
Age^2		–0.17	–0.04	-0.18		–0.08	–0.12
		(0.11)	(0.08)	(0.13)		(0.08)	(0.10)
Sender A in *Passive-Sender-B*				-0.18			
				(0.13)			
Sender B in *2-Sender*						–0.14	0.10
						(0.32)	(0.43)
Constant	1.09***	1.46*	1.51*	1.82*	1.62***	2.01***	0.78
	(0.06)	(0.74)	(0.89)	(1.03)	(0.18)	(0.60)	(1.60)
Observations	35	35	35	56	69	69	69
Clusters	35	35	35	56	37	37	37
F-statistic	22.14	19.65	17.77	8.01	13.11	3.79	3.09
R^2	0.53	0.76	0.88	0.63	0.12	0.21	0.26

regressions III and IV (*2-Sender*), the dependent variable is the senders' experienced anger. In regressions V and VI (*1-Sender*) as well as regressions VII and VIII (*2-Sender*), the dependent variable is the senders' experienced shame. While regressions I, III, V and VI include independent variables that indicate whether Message I or II was sent and whether Message II was dishonest (instead of honest), we added the interaction of normative views and second-order beliefs with the message that was sent to regressions II, IV, VI and VIII. All regressions include further controls for the antisocial premium, the interaction of the message sent with the senders' earnings from Message II, gender, age and, in case of the *2-Sender* game, for being Sender B.

Table A11 shows similar regressions as in Table A10. The only difference is the dependent variable, which is the senders' experienced happiness in regressions I and II (*1-Sender*) as well as in regressions III and IV (*2-Sender*). Furthermore, in regressions V and VI (*1-Sender*) as well as

Table A10. Determinants of other negative emotions

Note: Linear regressions of the senders' experienced negative emotions after observing the outcome of the game. Robust standard errors clustered on matched pairs (for senders in the *2-Sender* game) in parentheses. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	I	II	III	IV	V	VI	VII	VIII
Delivered Message II	3.10***	4.72**	0.63	2.41	0.72	2.28	−0.96**	−0.46*
	(0.99)	(2.25)	(0.48)	(1.49)	(0.54)	(1.53)	(0.42)	(1.25)
Delivered Message I × Normative views		0.15		−0.30***		−0.09		0.32
		(0.23)		(0.11)		(0.13)		(0.25)
Delivered Message II × Normative views		−1.00***		−0.41		0.02		0.19
		(0.39)		(0.37)		(0.20)		(0.19)
Delivered Message I × Second-order belief		0.02		0.37		0.19		0.46
		(0.60)		(0.61)		(0.37)		(1.07)
Delivered Message II × Second-order belief		1.85		−1.93		−2.36		0.37
		(3.08)		(1.54)		(1.88)		(0.91)
Deception	−0.43	−0.63	0.04	0.27	0.20	0.21	−0.91**	−1.02**
	(0.43)	(0.48)	(0.32)	(0.33)	(0.21)	(0.29)	(0.43)	(0.44)
Controls of individual characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	35	35	69	69	35	35	69	69
Clusters	35	35	37	37	35	35	37	37
F-statistic	7.31	7.35	1.61	9.26	0.56	0.71	3.26	2.73
R^2	0.64	0.77	0.28	0.39	0.27	0.31	0.32	0.35

regressions VII and VIII (*2-Sender*), the dependent variable is the senders' experienced gratitude. As in the case of guilt, these positive emotions refer to the moment senders learned the outcomes of all players and were elicited on a seven-point Likert scale ranging from 1 to 7 after the game was played.

Table A12 shows the linear regressions used to evaluate whether the treatment differences in the senders' experienced guilt after sending an antisocial message, that is followed by the receiver, are statistically significant between the *1-Sender* game and the *2-Sender* game in regression I and between the *1-Sender* game and the *Passive-Sender-B* game in regression II.

Table A11. Determinants of other positive emotions

Note: Linear regressions of the senders' experienced positive emotions after observing the outcome of the game. Robust standard errors clustered on matched pairs (for senders in the *2-Sender* game) in parentheses. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	I	II	III	IV	V	VI	VII	VIII
Delivered Message II	–0.91	–0.94	1.69***	1.14	0.09	–0.17	1.65***	0.79
	(0.96)	(2.28)	(0.42)	(1.50)	(0.89)	(3.48)	(0.52)	(1.82)
Delivered Message I × Normative views		0.20		–0.07		–0.27		–0.12
		(0.42)		(0.34)		(0.46)		(0.33)
Delivered Message II × Normative views		0.26		–0.34*		0.17		–0.10
		(0.19)		(0.18)		(0.35)		(0.29)
Delivered Message I × Second-order belief		0.58		0.75		–0.79		1.34
		(1.06)		(1.11)		(1.24)		(0.95)
Delivered Message II × Second-order belief		0.11		2.24*		–1.59		1.87
		(1.83)		(1.21)		(3.39)		(1.50)
Deception	0.29	0.18	0.19	0.24	0.85	1.06	–0.27	–0.19
	(0.48)	(0.48)	(0.50)	(0.56)	(0.60)	(0.71)	(0.52)	(0.56)
Controls of individual characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	35	35	69	69	35	35	69	69
Clusters	35	35	37	37	35	35	37	37
F-statistic	2.29	2.15	3.96	4.06	2.45	1.72	3.73	5.03
R^2	0.33	0.36	0.27	0.31	0.27	0.30	0.32	0.35

Table A12. Differences in experienced guilt levels

Note: Linear regressions of the senders' experienced guilt after observing the outcome of the game. Robust standard errors in parentheses. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

2-Sender	0.53***	0.53***
	(0.19)	(0.19)
Passive-Sender -B		0.04
		(0.14)
1-Sender × Message II	3.14***	3.14***
	(0.66)	(0.66)
2-Sender × Message II	1.06***	1.06***
	(0.29)	(0.29)
Passive-Sender-B × Message II		2.82***
		(0.49)
Constant	1.09***	1.09***
	(0.06)	(0.06)
Observations	104	147
Clusters	72	94
F-statistic	23.00	20.33
R^2	0.35	0.35

References

Behnk, Sascha, Li Hao, and Ernesto Reuben. 2017. "Understanding Diffusion of Responsibility in Antisocial Behaviors." Working Paper. New York University Abu Dhabi.