Engel, Christoph; Zhurakhovska, Lilia

Working Paper

Harm on an innocent outsider as a lubricant of cooperation: An experiment

Preprints of the Max Planck Institute for Research on Collective Goods, No. 2012,02

Provided in Cooperation with:
Max Planck Institute for Research on Collective Goods

**Preprints of the
Max Planck Institute for
Research on Collective Goods
Bonn 2012/2**

Harm on an Innocent Outsider as a Lubricant of Cooperation

An Experiment

Christoph Engel

Lilia Zhurakhovska

Preprints of the
Max Planck Institute
for Research on Collective Goods

Bonn 2012/2

Harm on an Innocent Outsider as a Lubricant of Cooperation

An Experiment

Christoph Engel / Lilia Zhurakhovska

February 2012

Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn
http://www.coll.mpg.de

Harm on an Innocent Outsider as a Lubricant of Cooperation

An Experiment

Christoph Engel[*] / Lilia Zhurakhovska[**]

Abstract

If two players of a simultaneous symmetric one-shot prisoner's dilemma hold standard preferences, the fact that choosing the cooperative move imposes harm on a passive outsider is immaterial. Yet if participants hold social preferences, one might think that they are reticent to impose harm on the outsider. This is not what we find, however severe the externality. A within-subjects measure of reticence to impose harm does not explain cooperation. But the externality makes participants more pessimistic. However conditional on their beliefs participants are more, not less cooperative if cooperation entails harm on an outsider, again however severe the externality.

JEL: C72, C91, D03, H23

Keywords: Prisoner's Dilemma, Externality, Modified Dictator Game, Beliefs

[*] Max Planck Institute for Research on Collective Goods, corresponding author: engel@coll.mpg.de

[**] Max Planck Institute for Research on Collective Goods, zhurakhovska@coll.mpg.de

1. Introduction

At Sunday school moral rules are clear. Morally, imposing harm on an innocent outsider is bad. Reproach is even stronger if the harm does not correspond to a direct benefit for insiders. Such action is not even selfish; it is purely spiteful. Now what if imposing harm does not benefit insiders individually, but rather benefits them jointly? This is the case if insiders face a dilemma, and if they impose harm on a bystander whenever at least one of them cooperates. In such a situation, the moral balance becomes more complicated. Resisting the temptation to exploit one's counterpart is usually regarded as unselfish and thereby morally desirable. But if the price for being social with an insider is harm on an outsider, the actor faces a choice between two morally condemned acts: selfishness and spite. Yet if spite is indeed worse than selfishness, and if at least some actors are guided by (these) moral principles, knowing that cooperation inflicts harm on a bystander should reduce cooperation.

Now the world is not Sunday school. But student subjects are usually expected to be rather social when they participate in lab experiments. After all, not more than a few dollars are at stake, and one would hope that, in terms of morality, students are a rather positive selection. When testing the power of the morally grounded reticence to impose harm on an innocent outsider in a lab experiment, we were therefore very surprised that widespread moral intuitions get it wrong. We tested participants on a one-shot symmetric prisoner's dilemma. In the treatment, insiders impose harm on a third, passive participant whenever at least one of them cooperates. Using the strategy method, we vary the degree of harm. However severe the harm, we do not find a significant difference between the baseline and the treatment.

The results are even more striking if we control for beliefs. As one would expect given the widespread moral norm against harming innocent victims, insiders are more skeptical about the cooperativeness of other insiders if cooperation imposes harm on the victim. Yet conditional on their more skeptical beliefs, they cooperate significantly more, however severe the harm. Knowing that a bystander will suffer is a lubricant of cooperation.

In fact, this result carries a second surprise. If both players of a symmetric one-shot prisoner's dilemma game hold standard preferences, both players defecting prescribes the unique equilibrium. As has been shown long ago, to a remarkable degree this prediction is violated in the lab (see already Rapoport and Chammah 1965). Some people just cooperate because of altruism, but most of the cooperation can be explained by conditional cooperation (Fischbacher, Gächter et al. 2001**;** Fischbacher and Gächter 2010). People cooperate because they deem it quite likely that other participants will cooperate as well. This explanation implies that the more participants are optimistic about cooperativeness, the more they are likely to cooperate; in turn, the more they are pessimistic, the less likely they are to cooperate. The very implication is violated in our data. Our participants are less optimistic about the cooperativeness of others in the presence of harm on outsiders, but surprisingly, conditional on their beliefs, they cooperate even more than in the baseline. Apparently, beliefs about cooperativeness are not the only condition that determines cooperative behavior. Our data suggests that conditional cooperators are also motivated by the

desire to distance themselves from outsiders. Apparently a cognitive component (beliefs) and a motivational component (increasing relative payoff) interact.

In the field, the conflict between kindness at the interior and meanness at the exterior is not infrequent. Sometimes, being mean is the very purpose of cooperation, as in a military coalition or in a trade union. At other instances, the harm is more a side-effect which is deliberately taken into account. Those closer to the source of a river build a dam, knowing that this deprives those closer to the estuary of the benefits of the river. Or a municipality builds a landfill to keep garbage off its streets, knowing that this puts the groundwater of neighboring municipalities at risk.

The most obvious motivation of our paper, however, is oligopoly. Viewed from inside the supply side of the market, competition may be interpreted as a prisoner's dilemma. In this perspective, collusion is the equivalent of cooperation, competitive behavior is defection. Individually, each supplier is best off if the other suppliers are faithful to the cartel, and she undercuts the collusive price or, for that matter, surpasses her quota. Yet if they cooperate, suppliers impose a distributional loss on the demand side, and they generate a deadweight loss, to the detriment of society.

The remainder of the paper is organized as follows: Section 2 relates the paper to the existing literature. Section 3 introduces the design. Section 4 makes theoretical predictions. Section 5 presents and discusses the treatment effects. Section 6 exploits the post-experimental tests to generate explanations for these effects. Section 7 concludes.

2. Related Literature

The effects of externalities on passive outsiders have only rarely been studied. To the best of our knowledge, they have not been tested in a standard prisoner's dilemma. Güth and van Damme (1998) present an ultimatum game with an externality on an inactive third player who has no say. The proposer decides how to divide the pie between three players. The division is executed if and only if the responder accepts. Otherwise, all three players receive nothing. In this game, the outsider receives very little. If the responder only learns the fraction the proposer wants to give the outsider, proposers keep almost everything for themselves. In anticipation, responders are very likely to reject the (mostly unknown) offer. Bolton and Ockenfels (2010) study lottery choice tasks in which the actor's choice also influences the payoff of a non-acting second player. This induces participants to take larger risks, provided the safe option yields unequal payoffs. Abbink (2005) plays a two-person bribery game in which corruption negatively affects passive workers. He concludes that reciprocity between briber and official overrules concerns about distributive fairness towards other members of the society. Ellman and Pezanis-Christou (2010) study how a firm's organizational structure influences ethical behavior towards passive outsiders. A firm of two players decides on its production strategy, which influences a passive third player. They find that horizontally organized firms in which the firm's decision corresponds to the average of both individual decisions are less likely to harm the outsider than consensus-based firms or firms in which one of both members is the boss. There is a rich experimental literature

on oligopoly (see the meta-study by Engel 2007); yet it does not focus on the fact that oligopoly is socially embedded.

Theories of conditional cooperation assume that there is heterogeneity in the willingness to cooperate in a dilemma. While some cooperate, others do not, and the conditional cooperator conditions her choices on information about cooperativeness. Ambrus and Pathak (2010) used a trust game to pre-classify participants by their cooperativeness. They composed groups of three cooperative and one selfish, or of three selfish and one cooperative player. Group composition had a pronounced effect on contributions to a public good. The most direct test of conditional cooperation stems from Fischbacher, Gächter et al. (2001**;** 2010). Using the strategy method (Selten 1967), they had participants to a linear public good make two choices: one unconditional choice, and one choice conditional on a complete table of other participants' contributions. For one participant, the first choice would be replaced by the contribution table, with the cell implemented that corresponded to the mean contribution of the remaining group members. The majority of participants conditioned their (second) choice on this information, but there was pronounced heterogeneity. The same procedure is used by Kocher, Cherry et al. (2008) in Austria and Japan, and by Herrmann and Thöni (2009) in Russia. Keser and van Winden (2000) compare a linear public good played in partner and in stranger design, and explain the difference by the tendency of participants to adjust behavior in the direction of the average behavior of the remaining group members, which they interpret as conditional cooperation. Croson, Fatas et al. (2005) have participants play a linear public good and show in regression analysis that the contributions of the remaining three group members in the previous period explain contributions of the fourth member now. This they interpret as a sign of conditional cooperation. By the same token, Frey and Meier (2004) give students different information about the fraction of students that have given to a charity in earlier terms, and find that giving is sensitive to this information.

3. Design

In our experiment, we have a *Baseline* with neither externalities nor sanctions and a treatment with negative *Externalities*. We deliberately avoid a market frame. This not only makes sure that our results are not driven by the frame. It is also necessary to isolate the effects of externalities. In a market setting, from their world knowledge subjects would know that collusion is illegal and might be motivated by this social and legal norm, rather than by their reticence to impose harm.

a) Baseline

Our baseline is a standard symmetric two-person-two-choices prisoner's dilemma, as in Table 1. If both players cooperate, each of them earns 5€. If one cooperates and the other defects, the co-operator earns nothing, while the defector earns 10€. If both defect, each of them earns 2.45€.[1]

	C	D
C	5€, 5€	0€, 10€
D	10€, 0€	2.45€, 2.45€

Table 1

Payoff Matrix *Baseline*

Our choice of parameters is primarily driven by experimental concerns. We create the maximum difference between the sucker payoff 0 and the temptation payoff 10. That way, both the premium for beating one's opponent and the penalty for losing in competition are largest. By contrast, the payoff in case both players defect almost holds the middle between the reward for cooperation and the penalty for being outperformed. For this payoff, we deliberately have not chosen either extreme. If participants earn 0€ in case both defect, cooperation is no longer strictly dominated. Strictly speaking, the game is no longer a prisoner's dilemma. At the opposite extreme, the equilibrium is not affected. But if participants earn 5€ in case both defect, gains from cooperation are 0. The situation is no longer a dilemma.

In a stylized way, our game also captures a one-shot Bertrand market with constant marginal cost where two firms individually decide whether to set the collusive price (C) or to engage in a price war (D). If both engage in (tacit or explicit) collusion, both set the monopoly price and split the monopoly profit evenly. If only one of them starts a price war, it undercuts the collusive price by the smallest possible decrement. As is standard in the theoretical literature, in this interpretation of our design we assume the decrement to be infinitesimally small, which implies that the aggressive firm cashes in the entire monopoly profit, while the firm that is faithful to the cartel receives nothing. If both firms start fighting, they end up in the Nash equilibrium. The positive payoff in the case of joint defection requires a slightly richer model, for instance one with heterogeneous products.

In a repeated game, the effects of optimism and reticence to impose harm would be overshadowed by reputation effects. We therefore test our subjects on a one-shot game. That way, we also need not be concerned that players might take turns. There is no room for an equilibrium in iterations.

[1] To make sure that the *Baseline* and our treatments are fully comparable, in the *Baseline* we also tested our participants on 11 problems that differed by just one parameter. To that end, we varied the payoff in case both defected between 0 € and 5 €. Since we do not need the additional data for our research question, we do not report these results. They are available from the authors upon request.

b) Externalities

In the *Externalities* treatment, payoffs for the active players are as in the *Baseline*. Yet in this treatment, each group consists of three players. If at least one of the two active players cooperates, a third, inactive player suffers harm h€. In a stylized way, this player captures the detrimental effects cooperating firms impose on the opposite market side, and on society at large. Using the strategy method (Selten 1967), we vary $h \in [.3€,9.3€]$, in 10 equal steps of .9€. This makes for the following payoff matrix:

	C	D
C	5€, 5€, $-h$€	0€, 10€, $-h$€
D	10€, 0€, $-h$€	2.45€, 2.45€, 0€

Table 2

Payoff Matrix *Externalities*

In the oligopoly interpretation of our design, this manipulation is meant to capture the loss in consumer welfare inherent in anticompetitive behavior. As in the field, this harm is not confined to the case of successful collusion. It also results if one firm sets the collusive price or quantity, while the other infinitesimally undercuts. Therefore, in the experiment, we do not confine harm to the situation where both active players cooperate. We impose the same harm if one cooperates while the other defects. We normalize harm to zero if both active players defect. Factor h thus captures the additional harm resulting from anticompetitive behavior. All participants made all choices in one of the randomly assigned roles of player A or B, and were after the experiment randomly matched with another participant. All problems were presented simultaneously on one computer screen. At the end of the experiment, one situation was chosen at random. We only gave feedback after the entire experiment was over.

c) Procedures

The experiment was run at the University of Bonn in May 2010 with a computerized interaction using z-Tree (Fischbacher 2007). ORSEE (Greiner 2004) was used to invite subjects from a subject pool of approximately 3500 subjects. Each subject played in one of the four treatments and no subject played in more than one. We collected 48 independent observations in both treatments; in the *Externalities* treatment, we also invited 24 inactive players, randomly assigned to be the potential targets of externalities. Subjects were on average 24.04 years old (range 17-50). 58.33% were female. They held various majors.[2] Each session lasted about one and a half hours. There was no show-up fee, but participants were guaranteed a minimum payoff of 5€.[3] Subjects earned on average 10.91€ (equivalent to 13.66$ on the last day of the experiment, range 5€-

[2] 22.08% lawyers, 13.75% economists.

[3] This applied to participants who had a total of less than 5€ from the main experiment and all post-experimental tests, especially if they made losses.

25.85€). In the *Baseline*, they earned on average 9.84€; in *Externalities*, the average sum was 11.80€. These earnings partly stem from post-experimental tests, which we report below.

4. Predictions

Since our game is a one-shot prisoner's dilemma, money-maximizing agents defect in the *Baseline*.

Empirically, many experimental subjects have been found to be conditional cooperators (Fischbacher, Gächter et al. 2001**;** Fischbacher and Gächter 2010). Pure conditional cooperators (at least weakly) prefer cooperation over defection if they expect their counterpart to cooperate with certainty. This implies that they resist the temptation to exploit their counterpart. If conditional cooperators are perfectly optimistic, they do not expect to run a risk. Consequently, in the *Baseline*, perfectly optimistic conditional cooperators cooperate.

In line with previous experiments, we expect conditional cooperation to be more prevalent than outright selfishness. Yet we expect participants to be less than perfectly optimistic. If their beliefs make them less optimistic, conditional cooperators run the risk of not getting gains from cooperation. If they are neutral to risk and losses, they compare the expected payoff of cooperation with the expected payoff of defection. If they are pure conditional cooperators in the sense of not desiring gains from exploitation, they discount gains from cooperation by their subjective degree of pessimism, and compare them with the minimum payoff in case they defect. Hence, for such actors, the size of this outside option matters. Cooperation is the less likely, the smaller the difference is between the outside option and gains from cooperation.

Cooperation becomes even less likely if an actor is an imperfect conditional cooperator, meaning that she strives to outperform her counterpart, if only slightly (Fischbacher and Gächter 2010); if she is averse against the risk of not getting gains from cooperation since her counterpart defects; if she dreads losing the outside option since she is exploited by her counterpart (Tversky and Kahneman 1992). If these personality traits combine, the dampening effect on cooperation multiplies.

If this actor defects while the other actor cooperates, two effects combine. Payoffs are unequal, with an advantage for the defecting actor (as modelled in Fehr and Schmidt 1999**;** Bolton and Ockenfels 2000). If the first actor expects the second to cooperate, she also violates the second actor's expectation of reciprocal action (as modelled in Rabin 1993**;** Dufwenberg and Kirchsteiger 2004). The reciprocity motive is not affected by adding a third player in treatment *Externalities*. Since the third player is inactive, she has no chance to reciprocate kind or unkind behavior. By contrast, in *Externalities* the inequity balance is more complicated. If both active players defect, they are symmetrically favored with respect to the inactive player. If both cooperate, they are favored even more. If one defects while the other cooperates, the defecting one is

strongly favored in comparison with both other players, while the cooperating one has a payoff of 0€, and the inactive player incurs a loss of –h€.

This line of argument, however, neglects that in case both active players defect, the payoff difference in comparison with the inactive players "is not their fault". Actually if they want to be kind to the inactive player, defecting is the best thing both can do. In situations that are structurally similar to the one tested here, it has been shown that intentions matter in the assessment of fairness (Falk, Fehr et al. 2008). Taking this into account, the *Externalities* treatment exposes active players to a conflict between fairness with the inactive player (calling for both defecting) and the motives behind conditional cooperation (calling for cooperation, provided the player is sufficiently optimistic about cooperativeness in this population). However, defection has a double dividend in this game: the defecting active player for herself at least secures the payoff she expects if both players defect cell, and she does the best she can to protect the inactive player from harm. The effect should be the stronger the more severe the harm on the outsider is. We therefore predict

Hypothesis: In *Externalities*, there is less cooperation than in the *Baseline*.

5. Treatment Effect

a) Baseline

The *Baseline* exposes participants to a standard prisoner's dilemma (with no externality on outsiders). The purpose of the baseline is to provide us with a benchmark. While the majority defected, 43.75% of our participants were willing to take the risk of cooperation.

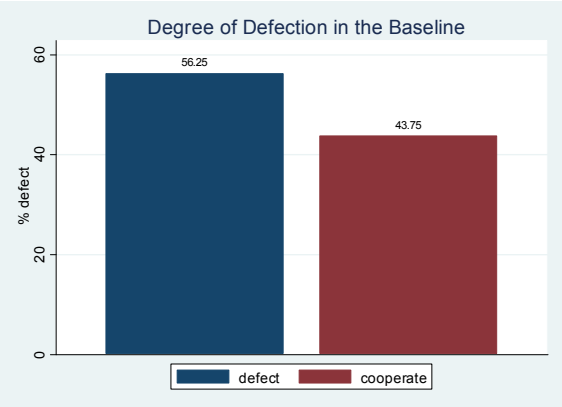

Figure 1

Degree of Defection in the Baseline

b) Externalities

In *Externalities*, the cooperation dividend (2.55€) implies that participants approximately double the outside payoff (2.45€). If they defect, participants have a chance to get a full 10€. If both cooperate, they impose considerable harm on the third player. Using the strategy method (Selten 1967), we varied harm, in equal steps of 0.90€, from -.30€ to - 9.30€. Those who had the bad fortune of being outside players lost a considerable amount of money (5 players lost 6.60€, 3 lost 4.80€, 5 lost 2.10€).[4] 13 of 24 outside players incurred losses. Figure 2 summarizes defection rates per game. Cooperation is pronounced. Even if they impose a loss of 9.30€ on outsiders, 33.33% of active participants still cooperate. The greater the harm, the less cooperation there is.[5]

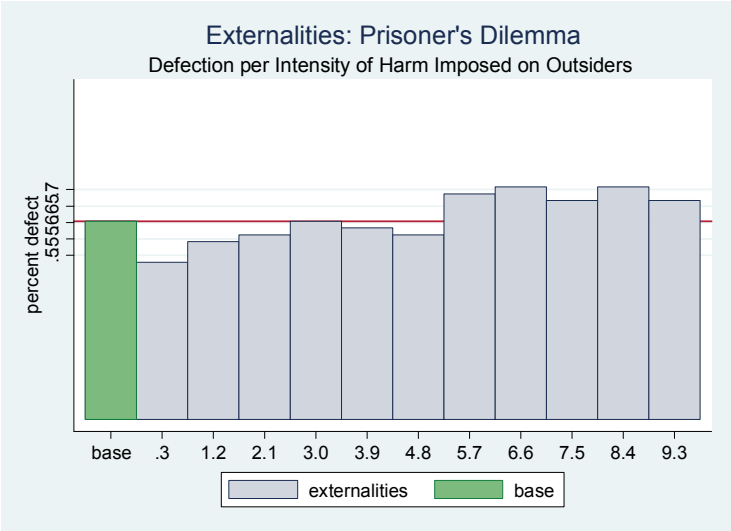

Figure 2

Externalities: Defection Rate per Game

x-axis: harm imposed on outsider (in €)

the horizontal line is at the level of defection in the *Baseline*

Descriptively, there is less cooperation than in the baseline with harm of 5.70€ or more. With smaller harm, descriptively there is even more cooperation than in the baseline. Yet Fisher's exact tests comparing the degree of cooperation in each of the 11 *Externality* games with the *Base-*

---

4 If they did not earn enough money in the remaining parts of the experiment, such participants received the minimum payoff of 5€.

5 OLS, explaining mean cooperation rate with level of harm, N = 11, coef .022, p < .001, cons .507, p < .001. We get the same result if we run a panel logit model, regressing individual choices for all 11 problems on levels of harm, N = 528, coef .240, p < .001, cons .418, p = .564. In this regression we work with $y_{ih} = \begin{cases} 1 & if \quad y^*_{ih} > 0 \\ 0 & if \quad y^*_{ih} \le 0 \end{cases}$, where y^*_{ih} is a latent variable defined over levels of harm h, nested in individuals i. The latent variable is a panel model, with $y^*_{ih} = \beta_1 + h\beta_2 + \nu_i + \varepsilon_{ih}$. We thus estimate the effect of the level of harm h, and include a subject-specific error term ν_i, which we assume to be unrelated with h and residual error ε_{ih}.

line are all insignificant. Hence we refute our hypothesis. Participants do not cooperate less if they know that cooperation imposes harm on outsiders.

Result 1: In a two-person simultaneous symmetric prisoner's dilemma, active players do not cooperate less if this imposes harm on an outsider.

6. Potential and Actual Driving Forces

a) Reticence to Impose Harm

We had expected that there would be less cooperation in *Externalities* since players might be reticent to impose harm on innocent outsiders. We have not found a significant difference between the *Baseline* and *Externalities*. To test whether the reticence to impose harm explains choices in a prisoner's dilemma with outsiders, after they have played the prisoner's dilemma, we tested our participants on a variant of the dictator game. We asked our subjects to choose between two situations: in situation 1, the proposer and her partner both got 5€. In situation 2, the proposer had a chance of $0 \le a \le 1$ to get 10€, and a chance of $1-a$ to get 5€, while the partner received nothing. The rules of the game were common knowledge. Again using the strategy method (Selten 1967), we varied $a \in [0,1]$, in equal steps of .1. We asked participants to make their choices for each of the 11 games. All participants made a choice in the role of the dictator, with random draws defining roles and matching participants, after the experiment. All problems were presented simultaneously on one computer screen. At the end of the experiment, one situation was chosen at random, and another random draw determined whether dictators made the high profit, provided they had chosen the lottery. We only gave feedback after the entire experiment was over. The game is as follows:

	Dictator	Recipient
Situation 1	5€	5€
Situation 2	a*10€+(1-a)*5€	0€

Table 3

Payoff Matrix Dictator Game Variant

We have our subjects choose between a lottery and a safe outcome, rather than between two safe outcomes, to maintain an element of risk. Both in the prisoner's dilemma and in this game, a player can make sure unilaterally that she will not fall below a modest payoff, while she must accept risk if she aims for a higher gain. In the prisoner's dilemma, if she defects, she at least earns the payoff for both players defecting (2.45€ in our case). Note that, in the prisoner's dilemma, there is both this risk (will the other player cooperate, which is a precondition for receiving 5€?) and a risk of incurring a loss (will the other player defect, which would reduce the payoff to zero?). Our design of the dictator game isolates the former motivational force. Whether the dictator gets a payoff higher than the sure 5€ hinges on a random draw (with stated probability). Yet the dictator can never fall below 5€, whether she is friendly with the recipient or not. Note

that the expected payoff of the active player is higher in situation 2 whenever a > 0, but the joint payoff of both players is higher in situation 1 as long as a < 1.

In this test, 46 of 48 active players in the prisoner's dilemma game (and the externalities treatment) were consistent, meaning that up until a certain probability of gaining 10€, they chose the equal split, while above that probability they always chose the lottery, which meant a payoff of 0 for the recipient. We can therefore work with switching points. Figure 3 summarizes the evidence. About a third of our participants maximized their payoff and seized the opportunity of a higher gain as soon as it was available. 7 participants were willing to spare the recipient, as long as the opportunity to get more for themselves was below 50%. 4 participants did not even injure the recipient if they were certain to have the double payoff.

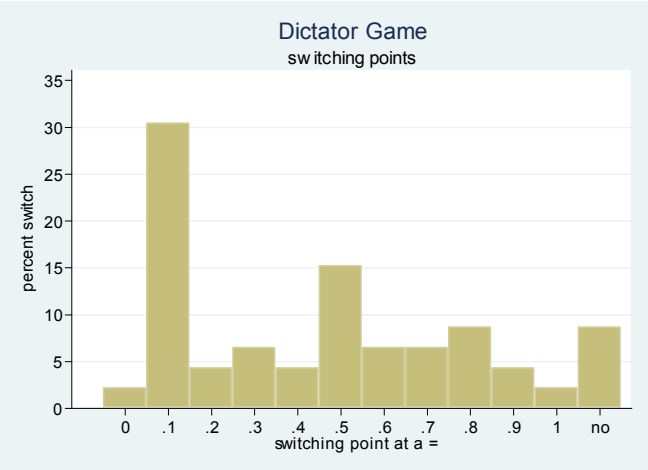

Figure 3

Dictator Game Variant

data from those 46 (of 48) active players in *Externalities* that were consistent in this test

the switching point is coded as *no* if a player never chooses situation 2

Information from the dictator game variant turns out almost completely uninformative for the prisoner's dilemma. If we regress choices in individual prisoner's dilemma problems, using logit models with a constant and heteroskedasticity-robust standard error, on the switching point in the dictator game, the regressor is weakly significant for the first problem, and insignificant for all remaining problems (Appendix Table 4). If we pool the data from the *Baseline* with each individual problem in *Externalities* and control for switching points in the dictator game, only in a single one of the 11 problems does the treatment dummy for *Externalities* in a logit model with a constant and heteroskedasticity-robust standard errors become weakly significant. This is the problem with the very large externality of 8.4€. The coefficient is positive and indicates that the probability of defection for this problem goes up by 17.58% to 73.83% (Appendix Table 5). We conclude:

Result 2: Reticence to impose harm does not explain the decision to cooperate in a simultaneous two-person prisoner's dilemma where cooperation imposes harm on an innocent outsider.

b) Optimism

Theoretically, the fact that we find a certain degree of cooperation at all levels of harm could result from the fact that participants are unconditionally cooperative. Yet earlier studies have normally only found a small number of participants who are willing to cooperate in a dilemma, whatever the remaining participants do. Many more are cooperative only conditional on the willingness of their experimental partners to cooperate as well (Fischbacher, Gächter et al. 2001; Fischbacher and Gächter 2010). For conditional cooperators, it is essential to estimate cooperativeness in the environment in which they happen to be. In repeated interaction, they may react to the experiences they have made in earlier rounds. Yet we test our subjects on one-shot games. Therefore all they have is their (home-grown) beliefs.

To understand the power of beliefs, after the main experiment, we elicit beliefs. We ask participants how many of the 24 participants of their session they think have chosen cooperation for one particular game. In *Externalities*, we do so for the case of $h = 6.6$. If participants get the number exactly right, they earn an additional 2€. If their estimate is within a range of +/- 2 around the true number, they earn an additional 1€. Feedback is given only after the entire experiment is over.

As Figure 4 shows, beliefs differ considerably across treatments.[6] Consequently, participants expect others to be sensitive to the fact that they impose considerable harm on an outsider. Nonetheless, as demonstrated earlier, their own choices are not significantly different from the *Baseline*.

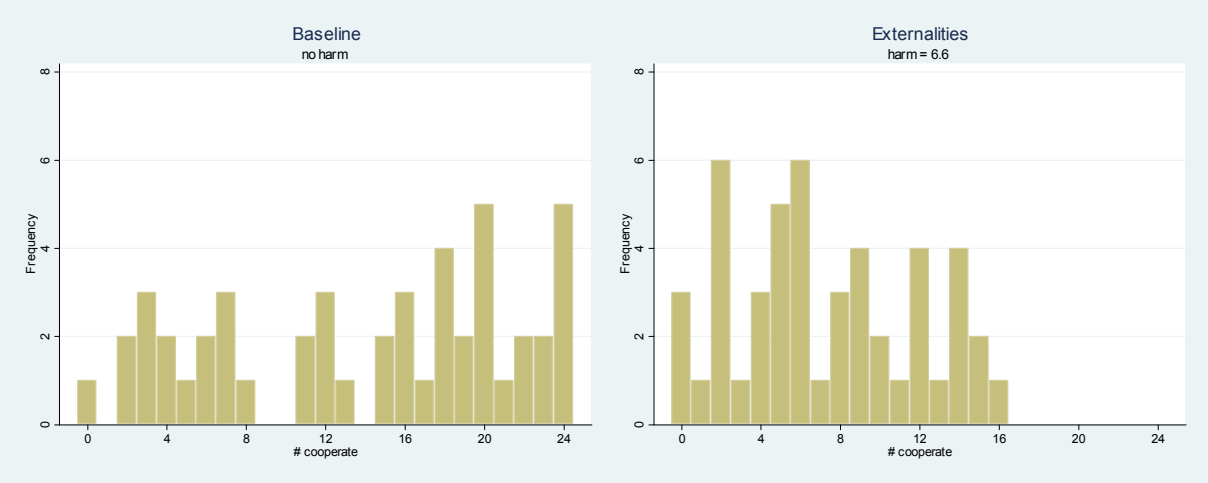

Figure 4

Beliefs

estimated number of cooperators, per treatment, in the indicated problem

[6] OLS, regressing the estimated number of cooperators (out of 24) on treatments. Treatment *Baseline* is reference category. Constant 14.042 ($p < .001$), *Externalities* -6.75 ($p < .001$), robust standard errors.

This already hints at the fact that, conditional on beliefs, participants are more willing to cooperate if cooperation entails harm on a passive outsider. This is indeed what we find if we pool data from the *Baseline* with data from each individual *Externalities* problem and control for beliefs, in a logit model with a constant and heteroskedasticity-robust standard errors. We now find a significant positive treatment effect for *all* 11 prisoner's dilemma problems (Appendix Table 6). This leads to the striking

Result 3: Conditional on their beliefs, however severe the harm they impose on an outsider, active players in a simultaneous two-person prisoner's dilemma cooperate more if cooperation is to the detriment of an outsider.

This result fits a recent finding from an experimental linear public good where contributions by active players to the public good either had a positive or a negative externality on passive bystanders. It turned out that contribution decisions were not driven by the direction of the externality, but by the comparison with bystander payoffs (Engel and Rockenbach 2011). In the present experiment, by the design of the *Externality* treatment, the outside player is always worse off. She can, at best, not lose, and will have her payoff reduced if at least one participant cooperates. Thereby cooperation pays a double dividend. Among the insiders, there is a chance to get gains from cooperation. If the other insider cooperates as well, both players also further distance themselves from the outsider. If not, relative payoffs depend on the size of the externality. Provided the externality is above 2.45€, even if a cooperator is exploited by the other insider she at least more strongly outperforms the outsider, compared with the only outcome she can enforce unilaterally and where both insiders earn 2.45€. Note that the benefit in terms of relative payoffs is the larger, the more pronounced the externality.

This finding also deepens our understanding of conditional cooperation. If the only condition for cooperation was the belief about cooperativeness, the fact that participants are significantly more pessimistic in *Externalities* should translate into less cooperation. Conditional on beliefs, there should not be a significant treatment effect. If there was a treatment effect, it should be explained by the fact that externalities induce more pessimism. Our findings are in clear opposition to these expectations. Instead, we find:

Result 4: If cooperation in a simultaneous two-person prisoner's dilemma entails a negative externality on an outsider, beliefs about cooperativeness and the decision to cooperate are negatively correlated.

Apparently, the belief about cooperativeness is not the only determinant of cooperation. It is in competition with the desire to distance oneself more strongly from bystander payoffs. This finding fits the result by (Neugebauer, Perote et al. 2009**;** Fischbacher and Gächter 2010). They have shown that most participants who are in principle willing to cooperate nonetheless desire to have a higher payoff than other active players. Through adding a third passive player, our design gives even more scope for payoff comparisons.

7. Conclusions

From the perspective of basic research, our endeavor has been successful. We have a highly surprising finding: if cooperation imposes harm on an innocent outsider, this does not make cooperation less likely in a symmetric one-shot-two-person prisoner's dilemma. More interestingly even: participants believe that the externality makes it less likely that their anonymous counterparts will cooperate, but conditional on their belief they react by a significantly higher willingness themselves to cooperate . This finding also qualifies the conventional understanding of conditional cooperation. Participants who are willing to cooperate in the first place not only condition their decision on information or beliefs about the willingness of other active players to contribute as well. This cognitive determinant is complemented by a motivational component. It results from the desire to outperform their peers.

From a policy perspective, our findings are less welcome news. Of course, industrial organization scholars always had second thoughts when analyzing competition as a stage game of profit-maximizing actors. The prediction of the Bertrand model (with homogenous goods) seemed too good to be true (see, e.g., the discussion in Tirole 1988**:** chapter 5). They were skeptical that the mere structure of the game would suffice to deter collusion. Yet our experiment was motivated by the hope that, at least, the fact that the suppliers' dilemma is embedded in a market would mitigate the otherwise pronounced ability to overcome the dilemma. As our results show, this hope is not well founded. Antitrust has reason to dread the willingness of suppliers to incur the risk of cooperation.

The latter effect, of course, requires that insiders compare themselves to outsiders. In the lab, this comparison is induced by the design of the experiment. In the field, insiders may not (always) consider themselves to be in the same boat as outsiders, which, in policy terms, would still imply that insiders collude if ever they can. At any rate, given our findings, antitrust has no reason to expect that reticence to impose harm on those at the opposite side of the market alleviates the cartel problem.

References

ABBINK, KLAUS (2005). Fair Salaries and the Moral Costs of Corruption. Advances in Cognitive Economics. B. N. Kokinov. Sofia, NBU Press.

AMBRUS, ATTILA and PARAG A. PATHAK (2010). "Cooperation over Finite Horizons. A Theory and Experiments." Journal of Public Economics **95**: 500-512.

BOLTON, GARY E. and AXEL OCKENFELS (2000). "ERC: A Theory of Equity, Reciprocity and Competition." American Economic Review **90**: 166-193.

BOLTON, GARY E. and AXEL OCKENFELS (2010). "Betrayal Aversion. Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. Comment." American Economic Review **100**: 628-633.

CROSON, RACHEL T.A., ENRIQUE FATAS, et al. (2005). "Reciprocity, Matching and Conditional Cooperation in Two Public Goods Games." Economics Letters **87**: 95-101.

DUFWENBERG, MARTIN and GEORG KIRCHSTEIGER (2004). "A Theory of Sequential Reciprocity." Games and Economic Behavior **47**: 268-298.

ELLMAN, MATTHEW and PAUL PEZANIS-CHRISTOU (2010). "Organisational Structure, Communication and Group Ethics." American Economic Review **100**: ***.

ENGEL, CHRISTOPH (2007). "How Much Collusion? A Meta-Analysis on Oligopoly Experiments." Journal of Competition Law and Economics **3**: 491-549.

ENGEL, CHRISTOPH and BETTINA ROCKENBACH (2011). We Are Not Alone. The Impact of Externalities on Public Good Provision.

FALK, ARMIN, ERNST FEHR, et al. (2008). "Testing Theories of Fairness - Intentions Matter." Games and Economic Behavior **62**: 287-303.

FEHR, ERNST and KLAUS M. SCHMIDT (1999). "A Theory of Fairness, Competition, and Cooperation." Quarterly Journal of Economics **114**: 817-868.

FISCHBACHER, URS (2007). "z-Tree. Zurich Toolbox for Ready-made Economic Experiments." Experimental Economics **10**: 171-178.

FISCHBACHER, URS and SIMON GÄCHTER (2010). "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Good Experiments." American Economic Review **100**: 541-556.

FISCHBACHER, URS, SIMON GÄCHTER, et al. (2001). "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." Economics Letters **71**: 397-404.

FREY, BRUNO and STEPHAN MEIER (2004). "Social Comparisons and Pro-social Behavior: Testing “Conditional Cooperation” in a Field Experiment." American Economic Review **94**: 1717-1722.

GREINER, BEN (2004). An Online Recruiting System for Economic Experiments. Forschung und wissenschaftliches Rechnen 2003. K. Kremer and V. Macho. Göttingen**:** 79-93.

GÜTH, WERNER and ERIC VAN DAMME (1998). "Information, Strategic Behavior, and Fairness in Ultimatum Bargaining. An Experimental Study." Journal of Mathematical Psychology **42**: 227-247.

HERRMANN, BENEDIKT and CHRISTIAN THÖNI (2009). "Measuring Conditional Cooperation. A Replication Study in Russia." Experimental Economics **12**(1): 87-92.

KESER, CLAUDIA and FRANS VAN WINDEN (2000). "Conditional Cooperation and Voluntary Contributions to Public Goods." Scandinavian Journal of Economics **102**: 23-39.

KOCHER, MARTIN, TODD L. CHERRY, et al. (2008). "Conditional Cooperation on Three Continents." Economics Letters **101**(3): 175-178.

NEUGEBAUER, TIBOR, JAVIER PEROTE, et al. (2009). "Selfish-biased Conditional Cooperation. On the Decline of Contributions in Repeated Public Goods Experiments." Journal of Economic Psychology **30**(1): 52-60.

RABIN, MATTHEW (1993). "Incorporating Fairness into Game Theory and Economics." American Economic Review **83**: 1281-1302.

RAPOPORT, ANATOL and ALBERT M. CHAMMAH (1965). Prisoner's Dilemma. A Study in Conflict and Cooperation. Ann Arbor,, University of Michigan Press.

SELTEN, REINHARD (1967). Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. Beiträge zur experimentellen Wirtschaftsforschung. E. Sauermann. Tübingen, Mohr**:** 136-168.

TIROLE, JEAN (1988). The Theory of Industrial Organization. Cambridge, Mass., MIT Press.

TVERSKY, AMOS and DANIEL KAHNEMAN (1992). "Advances in Prospect Theory. Cumulative Representation of Uncertainty." Journal of Risk and Uncertainty **5**: 297-323.

Appendix

I. Instructions

The Instructions for the *baseline* and the *externalities treatment* differ only in Part 1. The rest is identical. Therefore we report first the full instructions of the baseline treatment and afterwards only part 1 of the treatment.

a. Baseline

Welcome to our experiment. Please remain quiet and do not talk to the other participants during the experiment. If you have any questions, please give us a signal. We will answer your queries individually.

Course of Events

The experiment is divided into four parts.[7] We will distribute separate instructions for each of the four parts of the experiment. Please read these instructions carefully and make your decisions only after taking an appropriate amount of time to reflect on the situations, and after we have fully answered any questions you may have. Only when all participants have decided will we move on to the next part of the experiment. All of your decisions will be treated anonymously.

Your Payoff

At the end of the experiment, we will give you your payoff in cash. Each of you will receive the earnings resulting from the decisions you will have made in the course of the experiment. It is possible to make a loss in one part of the experiment. These losses will be subtracted from the earnings in the other parts.

Thus:

Total payment =

+ Earnings from Part 1
+ Earnings from Part 1a
+ Earnings from Part 2
+ Earnings from Part 3
+ Earnings from Part 4
(min. 5€)

[7] Part 3 was a post-experimental test of risk and loss aversion. Part 4 was a post-experimental test of social value orientation. We use neither test for this paper, and therefore also do not reproduce that part of the instructions. They are available from the authors upon request.

In Part 2, however, losses are possible, too. Should you incur losses, these will be deducted from your earnings from Part 1, Part 3, or Part 4. (The possibility of losses in Part 2 is limited, however; you will definitely receive a total payment that is on the plus side of the balance.) If you earn on the whole less than 5€, you will get a minimum payment of 5€.

We will explain the details of how your payoff is made up for each of the four parts separately. In each of the four parts, possible payoffs are given in Euro, which is the currency you will be paid in.

Part 1

The basic idea of this part of the experiment is as follows: you are anonymously paired by us with another participant. You and the other participant will make a total of eleven decisions.

Only one pair of decisions will determine your payoff. This procedure is explained below.

We will show you eleven tables that look as follows:

		Type B	
		Above	*Below*
Type A	*Above*	5€, 5€	0€, 10€
	Below	10€, 0€	z€, z€

We will let you know at the start whether you are a Type A or a Type B participant. (You will probably notice that the payments given to both types are symmetrical; the distinction between Type A and Type B is solely for the purpose of explaining the experiment.)

The decisions *Above* or *Below* determine the payoffs to you and the other participant. In each of the four cells of the table, the figure on the left denotes A's profit, while the figure on the right denotes B's profit.

For instance, if Type A chooses the option *Above* and Type B chooses the option *Above*, then both receive a payment of 5€. If Type A chooses *Above* and Type B chooses *Below*, then Type A receives zero profit and Type B gets 10€. The same is valid for a *Below*/*Above* constellation. Finally, if Type A chooses *Below* and Type B chooses *Below*, then both receive a payment of z€.

What does the z stand for? z is varied in the following eleven tables; all other payments remain unchanged. You have to decide on all eleven tables (*Above* or *Below*). Please mark your decision by clicking on the appropriate box shown on your screen.

You will be free to address each of the eleven tables separately, making your decisions independently of the other tables. You can also make the same decision all the time. This is entirely up to you.

Please note, once again, that only one of the eleven decision pairs will be relevant for your payoff. We will choose one of the eleven tables at random at the end. Your decision for the table that is drawn by lot and the other participant's decision for the same table determine the payoff in this part of the experiment.

Let us first begin with some test questions. (The aim of these questions is merely to verify whether all participants have fully understood the instructions. Neither the questions nor the answers have anything to do with your final payment.) Then the screen on which your actual decisions are marked will appear.

Do you have any further questions?

Part 1a

This part of the experiment refers to the previous part where you made eleven decisions, "Above" or "Below". The number of participants who participated in this task will be presented to you on the screen. We ask you to estimate how many participants of the experiment selected "Above" for a particular Z (see the decision screen for detailed information). In case you make a precise estimation, you can gain 2€ in addition. If your estimation deviates by +/-2, you still gain 1€ in addition. Otherwise, you gain nothing in addition.

Part 2

This part of the experiment is as follows: one Type X participant has to decide between two situations (1 or 2). His decision influences his own payoff, and the payoff of one other randomly paired Type Y participant, as follows:

Situation 1: Type X receives a payoff, determined by lot, of 5€ or 10€, Type Y receives a payoff of zero Euro. The likelihood with which Type X either receives 5€ or 10€ is systematically varied in the following table. Type X must make a decision for each of the eleven constellations (a total of 11 decisions).

Situation 2 remains the same for all 11 constellations: Type X and Type Y both receive 5€.

In this part, all participants must initially make their decisions in the role of Type X.

We will proceed with the payoff as follows:

- The lot is drawn to determine whether your payments, following your own decisions, classify you as a Type X or a (passive) Type Y. We will draw one half of the group as Type X and the other as Type Y.
- The next draw pairs each Type Y participant with a Type X participant.
- Finally, the third draw determines one single payoff-relevant situation out of the total of eleven situations. Therefore, one out of the eleven decisions emerges as the basis for payoff. With a probability of ½, it will be your own decision, and with the same likelihood it will be another participant's decision.

Example for Part 3

	Profit	With likelihood of	
You	10€	30%	Situation 1
	5€	70%	
Other participant	0€	100%	
Your decision	1		
	2		
Both	5€	100%	Situation 2

As stated above, all participants will make eleven decisions of this kind. Please mark your decision by clicking on the appropriate box.

b. Externalities

Part 1

The basic idea of this part of the experiment is as follows: you are anonymously paired by us with two other participants. There exist Type A, Type B and Type C players. Type C is passive in that experiment. If you are not Type C, you and one other participant will make a total of eleven decisions.

Only one pair of decisions will determine your payoff. This procedure is explained below.

We will show you eleven tables that look as follows:

		Type B	
		Above	Below
Type A	Above	5€, 5€, -D€	0€, 10€, -D€
	Below	10€, 0€, -D€	2.45€, 2.45€, 0€

We will let you know at the start whether you are a Type A or a Type B participant. (You will probably notice that the payments given to both types are symmetrical; the distinction between Type A and Type B is solely for the purpose of explaining the experiment.)

The decisions Above or Below determine the payoffs to you and the other participants. In each of the four cells of the table, the figure on the left denotes A's profit, while the figure on the right denotes B's profit. Type C receives either –D€ or 0€, depending on the decisions of Type A and B.

For instance, if Type A chooses the option Above and Type B chooses the option Above, then both receive a payment of 5€ and Type C receives –D€. If Type A chooses Above and Type B chooses Below, then Type A receives zero profit, Type B gets 10€, and Type C receives –D€. The same is valid for a Below/Above constellation. Finally, if Type A chooses Below and Type B chooses Below, then both receive a payment of 2.45€ and Type C receives 0€.

What does the D stand for? D is varied in the following eleven tables. It is an absolute value that will be paid in €; all other payments remain unchanged. You have to decide on all eleven tables (Above or Below). Please mark your decision by clicking on the appropriate box shown on your screen.

You will be free to address each of the eleven tables separately, making your decisions independently of the other tables. You can also make the same decision all the time. This is entirely up to you.

Please note, once again, that only one of the eleven decision pairs will be relevant for your payoff. We will choose one of the eleven tables at random at the end. Your decision for the table that is drawn by lot and the other participant's decision for the same table determine the payoff in this part of the experiment.

Let us first begin with some test questions. (The aim of these questions is merely to verify whether all participants have fully understood the instructions. Neither the questions nor the answers have anything to do with your final payment.) Then the screen on which your actual decisions are marked will appear.

Do you have any further questions?

For Online Appendix

II. Supplementary Data Analysis

level of harm	.3	1.2	2.1	3	3.9	4.8	5.7	6.6	7.5	8.4	9.3
dictator game switching point	-.168 (.091)	.044 (.611)	-.137 (.144)	.014 (.875)	-.094 (.297)	.039 (.654)	-.046 (.628)	.067 (.521)	-.013 (.888)	.146 (.211)	.048 (.632)
Cons	.677 (.191)	.062 (.902)	.909 (.082)	.469 (.360)	.797 (.127)	.171 (.734)	1.044 (.063)	.745 (.183)	.788 (.149)	.426 (.449)	.509 (.349)
N	46	46	46	46	46	46	46	46	46	46	46

Table 4

Explaining Choices in Individual *Externalities* Problems

with Switching Point in Dictator Game

logit, with robust standard errors, p-values in parentheses

level of harm	.3	1.2	2.1	3	3.9	4.8	5.7	6.6	7.5	8.4	9.3
Externalities	-.546 (.223)	-.056 (.895)	-.154 (.727)	.203 (.638)	-.037 (.932)	.031 (.942)	.475 (.285)	.729 (.104)	.385 (.379)	.753 (.089)	.409 (.345)
dictator game switching point	-.161 (.015)	-.062 (.301)	-.147 (.023)	.077 (.203)	-.127 (.044)	-.064 (.285)	-.107 (.089)	-.064 (.303)	-.091 (.141)	-.037 (.551)	-.065 (.289)
Cons	1.194 (.018)	.607 (.173)	1.110 (.023)	.697 (.124)	.990 (.038)	.622 (.165)	.872 (.064)	.622 (.174)	.781 (.091)	.464 (.305)	.627 (.167)
N	94	94	94	94	94	94	94	94	94	94	94

Table 5

Comparing *Baseline* with Individual *Externalities* Problems,

Controlling for Switching Point in Dictator Game

logit, with robust standard errors, p-values in parentheses

level of harm	.3	1.2	2.1	3	3.9	4.8	5.7	6.6	7.5	8.4	9.3
ext	-2.205 (0.001)	-2.025 (0.002)	-1.738 (0.004)	-1.965 (0.004)	-2.056 (0.004)	-2.091 (0.002)	-1.641 (0.016)	-1.418 (0.026)	-1.747 (0.01)	-1.602 (0.02)	-1.947 (0.008)
belief	-0.233 (<.001)	-0.248 (<.001)	-0.228 (<.001)	-0.28 (<.001)	-0.277 (<.001)	-0.268 (<.001)	-0.302 (<.001)	-0.291 (<.001)	-0.299 (<.001)	-0.315 (<.001)	-0.324 (<.001)
cons	3.779 (<.001)	4.028 (<.001)	3.696 (<.001)	4.555 (<.001)	4.502 (<.001)	4.354 (<.001)	4.927 (<.001)	4.736 (<.001)	4.869 (<.001)	5.147 (<.001)	5.288 (<.001)

Table 6

Comparing *Baseline* with Individual *Externalities* Problems,

Controlling for Beliefs

logit, with robust standard errors, p-values in parentheses