How Norms Guide Behavior in an Uncertain World
An Experimental Law & Economics Perspective

Dissertation
zur Erlangung des akademischen Grades
doctor rerum politicarum
(Dr. rer. pol.)

vorgelegt dem
Rat der Wirtschaftswissenschaftlichen Fakultät
der Friedrich-Schiller-Universität Jena

am 17. Oktober 2012

von M.Sc. Michael Kurschilgen

geboren am 16.07.1982 in Bonn

Gutachter

Prof. Dr. Christoph Engel, Max-Planck-Institut zur Erforschung von Gemeinschaftsgütern, Bonn

Prof. Dr. Werner Güth, Max-Planck-Institut für Ökonomik, Jena

Prof. Dr. Oliver Kirchkamp, Friedrich-Schiller-Universität Jena

Datum der Verteidigung: 23. Januar 2013

Für hilfreiche Ratschläge, Kommentare und Kritik an früheren Versionen der Arbeit danke ich Christoph Engel, Werner Güth, Oliver Kirchkamp, Sebastian Kube, Shachar Kariv, Matthew Rabin, George Akerlof, Aniol Llorente-Saguer, Christian Traxler, Sebastian Schneiders, Sebastian Siegloch, Carlos Kurschilgen und Laura Hennige.

A Ut & Yrus

Content

Introduction

Life is full of uncertainty. For many of the decisions we take, we ignore the exact consequences of choosing one way or the other. For example, when we face a red traffic light we often do not know how dangerous the crossing is, and whether the police might be watching. When we decide to buy a house, we do not know whether we will get along well with our neighbors, and how the market for that house will be in case we want to resell it. When we consider giving money to a charity, we cannot be sure they really spend the money as they claim, and whether those proposed measures actually help the people in need.

In this uncertain world, norms help us take decisions by providing guidance. Norms may have different sources. They can be institutionally enacted by a legislator, common behavioral patterns within a social group, and individually held moral standards. We may for instance choose not to run a red light, simply because doing so is illegal; totally irrespective of the specific situation at a particular crossing. We may decide to buy a house because most of our friends and colleagues own houses. And we may give money to a charity simply because we believe that something needs to be done.

In four independent Chapters, I examine how norms guide our behavior in an uncertain world. Methodologically, I apply the toolbox of experimental economics to research questions that are on the interface between economics and law (especially Chapter 1 and Chapter 2), sociology (especially Chapter 3), and social psychology (especially Chapter 4).

The first Chapter looks into the market for copyrights. Imagine the following situation: An author has written a script for a movie and a producer is considering buying a license in order to make a film. At the time when they have to strike a deal, neither of the two parties knows to how much box office success this script might be able to contribute. The reason is that the distribution of box office revenues is highly skewed. Few movies make a lot of money. The vast majority of movies barely recover the production costs. So far, there are no reliable methods for predicting box office success accurately. In such a situation it is difficult for either party to determine its reservation price, and even more complicated to agree to a mutually acceptable

deal. The German legislator thinks that this requires regulation and stipulates that the seller of the copyright may ex post claim additional remuneration in case that the license fee agreed upon ex ante is grossly disproportionate to the proceeds from the work.

This so-called "bestseller paragraph" seems odd in at least two respects: First, it is unbalanced as it does not grant the buyer the right of an ex-post discount in case the ex-ante licence fee was disproportionately high. Second, it interferes with a voluntarily agreed, mutually beneficial contract. Together with Christoph Engel, I examine to what extent the provision restores fairness, as claimed in the legislative materials, and how it affects the number of deals closed as well as the average prices in the market.

We design an experimental game that captures the main features of copyright markets: A seller and a buyer negotiate a price for a commodity whose value is uncertain. Only after they have agreed on a purchase price, nature determines the commodity's true value. Both parties then have the opportunity to costly express their discontent to one another. One treatment pictures the market *without* and another *with* the essence of the German provision. In the *provision* treatment, buyer and seller can renegotiate the price after the draw of nature. In case they do not agree, a third person, the so-called umpire, determines the purchase price.

Under standard textbook assumptions the equilibrium of the game is not touched by the presence of the bestseller paragraph. Under behavioural assumptions we expect lower market prices but no effect on the number of deals closed. We further expect ex-post discontent to be distributed differently but not reduced on aggregate. The experimental results show that the provision leads to lower market prices and a higher number of deals closed. Apparently thus, the provision enhances ex-ante fairness in the sense that lower priced deals become more acceptable to sellers. It also increases ex-post fairness since buyers express less discontent while seller express as little as in the *baseline*.

Whilst the first Chapter is concerned with codified law, Chapter 2 shifts the focus to customary law. In most countries of the world, customary law is a recognised source of law. It is created by the behaviour and the will of those supposed to abide by it. Classic illustrations are international law, *lex mercatoria,* trade practice, and codes of conduct. These domains are characterized by substantial uncertainty concerning the

behaviour of the other players and whether a certain practice can already be regarded as sufficiently established to have binding character.

Together with Christoph Engel, I use a lab experiment to test the main claims made in the legal debate over customary law. Rational choice theorists claim that what looks like custom is nothing but self-interest. Positivists doubt that anything beyond consent assumes the force of law.

The experimental paradigm is a standard public goods game with and without sanctioning opportunities. We vary the extent to which participants are made aware of the binding force of custom and whether we elicit their normative expectations or not. Our experimental results suggest that the critics of customary law use an overly narrow and thus inappropriate conception of normativity. Norms do not only direct behaviour if norm violation is against the addressee's self-interest. On the other hand, the duty to abide by the law in force is not the exclusive motivating force either. Norms matter because they provide guidance. Most actors are most of the time willing to follow the norms prevailing in their context, or at least to be not too far off the mark, and most actors expect other actors to be thus guided. This leads to considerably higher contributions to the public good.

Based on these findings, we propose a new conceptualisation of customary law: Custom guides behavior as normative expectations and behavioral patterns coevolve. Customary law capitalizes on this more general social mechanism. If it is not backed up by sanctions, customary law is not more effective than mere custom in realigning individual action and social welfare. Sanctions and custom are substitutes if the emerging rule is not perceived to be grounded in law, and complements otherwise.

Chapter 3 leaves the domain of legal norms and shifts the focus to social norms. Social norms are crucial when individuals need to cooperate in order to achieve a better social outcome. Cooperation problems are at the heart of many everyday situations. For example, when it comes to protecting the environment, defending one's country, generating new knowledge, joining a political party, extending the infrastructure, or exploiting the opposite market side, agents face a social dilemma. Jointly they are best off if everyone contributes her fair share. But individually, free-riding on others' efforts yields the highest payoff.

Together with Christoph Engel and Sebastian Kube, I propose a very simple and light-handed mechanism for managing cooperativeness and test its effectiveness by using laboratory experiments. Our mechanism is based on an observation that has previously been made in public-goods experiments, namely that (initial) group composition and initial cooperation rates significantly affect the future development of cooperation rates in a particular group. A likely reason for this behavioral pattern is provided by the influential concept of reciprocity and conditional cooperation. If persons are sensitive to the behavior of other group members, those groups who cooperate little in the beginning will become even less cooperative over time, while groups with substantial cooperation in the beginning are able to sustain cooperation over time. This suggests that outsiders might moderate cooperation by controlling experiences. Our mechanism is even less invasive in that it is confined to the first impressions subjects happen to make. More precisely, the idea is to "manipulate" initial beliefs by providing participants with selective information about (un)cooperative behavior in other, unrelated, groups. We thus suggest to the participants a possible norm to follow.

The experimental paradigm is a public goods game with decentralized sanctions and counter-punishment opportunities. We choose this rich environment because it mimics several potentially important features available in natural environments. Moreover, it is sufficiently complex, so that conflicting behavioral norms might emerge – which allows us to test the effect of both favorable and unfavorable first impressions.

In the baseline, subjects play the game for 10 periods without any exogenous interventions. In two treatments, prior to making their first contribution decision in the game, we provide subjects with selected data from the baseline. Specifically, we show them graphs of the development of mean contributions over time from selected cooperative (resp. uncooperative) groups. The information that we give is unfavorable for cooperation in one treatment and favorable in the other.

Our experimental results show that contributions are considerably sensitive to such selective information. In fact, the selective information from unrelated groups appears to set a standard as to what type of behavior is expected to be *normal*. First impressions participants happen to make predict subsequent behavior. Our results, however, suggest an asymmetry in the strength of the reaction – which might pose a limit on the effectiveness of the mechanism in natural settings. People are particularly

sensitive to negative impressions as optimistic beliefs about others' behavior are much more fragile than pessimistic ones. These findings strongly suggest that people's behavior is indeed conditional on initial expectations of how others are likely to behave, and that these expectations are open to purposeful intervention.

In most everyday situations we know which norm applies and adapt our behavior automatically and unconsciously. But often, we face new situations and are uncertain what the *right* behavior is. We find guidance in norms. Absent legal norms, there are two distinct sources from which to deduce a norm. The first source is external. People watch how others behave and deduce the norm from what most people do. I call this a norm of conformity. Chapter 3 illustrates the (detrimental) effect of conformity in a social dilemma. On the other hand, the source of a norm can also be internal. People look at their own moral standards and deduce the norm from what they personally believe to be morally right. I call this a norm of morality. Chapter 2 hints at the beneficial effect of morality in a social dilemma. Conformity requires information about others' behavior. Morality requires self-reflection about one's own moral standards. This is the rationale for the two experimental treatments. Chapter 4 studies the effect of these two distinct normative sources on people's social preferences.

While the idea of self-reflection is rather strange to modern economics, it is not new to the discipline. In fact, it can be traced back to Adam Smith's Theory of Moral Sentiments, who in turn was inspired by the religious concept of conscience. More recently, the concept of identity utility has provided a rationale for explaining preference changes due to shifts of the relevant norms; be it unconsciously or via conscious self-reflection. This contrasts sharply with the predominant notion in modern economics that preferences are inherent and stable traits of the individual. Notably, the literature on social preferences acknowledges that different people might have different preferences for distributive justice but it does not allow the same people to vary their preferences from one situation to another as a result of a change in the relevant norms.

The experimental paradigm used in Chapter 4 is a modified dictator game (MDG). The MDG asks participants for their preferences in a 2x2 world: you are either richer or poorer than the other player, and you can either create income for the other player or destroy it. A MDG in a university lab is clearly not an everyday situation. It thus seems plausible to assume that participants have doubts about the normatively appropriate behavior, which is exactly where normative guidance becomes relevant.

In the experimental treatments, subjects are guided externally, by giving them information about representative behavior of previous participants, and internally, by asking them to make a moral judgment prior to their incentivized choice.

The experimental results suggest that whereas information has virtually no effect, self-reflection changes social preferences substantially. Information does have, however, a strong effect on people's moral judgments. Interestingly, since people differ with respect to their normative goals (social welfare vs. equality), self-reflection leads to more heterogeneity of preferences as subjects home-grown normative concerns are reinforced.

Chapter 4 is single-authored. Chapters 1, 2, and 3 have been written together with co-authors. My personal contributions to the co-authored Chapters are summarized in Table I.

Table I: Personal contribution to co-authored Chapters

	Chapter 1	Chapter 2	Chapter 3
Idea	Minor	Minor	Proportional
Experimental Design	Proportional	Proportional	Proportional
Hypotheses	Leading	Proportional	Proportional
Literature Review	Minor	Minor	Proportional
Data Collection	Leading	Leading	Leading
Data Analysis	Leading	Proportional	Leading
Writing	Proportional	Proportional	Proportional

Chapter 1

Fairness Ex Ante & Ex Post:
Experimentally Testing Ex Post Judicial Intervention into Blockbuster Deals

Christoph Engel and Michael Kurschilgen

Abstract

The market for copyrights is characterised by a highly skewed distribution of profits: very few movies, books and songs generate huge profits, whereas the great bulk barely manages to recover production cost. At the moment when the owner of intellectual property grants a licence ("ex ante"), neither party knows the true value of the traded commodity. A seemingly odd provision from German copyright law, the so-called "bestseller paragraph", stipulates that the seller of a licence has a legally enforceable right to a bonus in case the work ("ex post") turns out a blockbuster. We experimentally explore the effect of the provision on market prices, on the number of deals struck and on perceived fairness. Our results show that the provision leads to lower prices for copyrights. More copyrights trade. The buyers express less ex-post discontent.

Keywords: Copyrights, Fairness, Experiment

JEL: C91, D03, D63, D81, K12, O34

I. Introduction

Choosing the price of a copyright is difficult. The distribution of success is highly skewed in the media industries and success is highly unpredictable.

Take the example of the movie industry. In 2008, the most successful movie, *The Dark Knight*, in total gross earnings made more than $ 1 billion worldwide at the box office, whereas the least successful release from a studio, called *$ 9.99*, just made $ 800.[1] In stylised facts: 20 % of the films earn approximately 80 % of the revenue (De Vany and Walls 1996; Jedidi and Krider 1998:394; Collins, Hand et al. 2002). The majority of movies even generate real losses (De Vany and Walls 1999:298 provide an illustrative scatterplot).

The dynamics of film success have been explained by herding and information cascades (De Vany and Walls 1996). However, with the information available ex ante, it is extremely difficult to predict success. Even using regression coefficients from the past, and exploiting all the information available ex ante, gross mispredictions are frequent. For instance, the movie *3 ninjas* actually made $ 308,000, while the regression would have predicted more than $ 10 million. By contrast, for *There's Something About Mary*, the regression predicted less than $ 2 million, while the movie actually made more than $ 175 million (Simonoff and Sparrow 2000:Table 2).[2]

Imagine that an author has written a script for a movie and there is a producer who considers buying a license in order to make a film. At the time when the producer has to decide how much to pay, neither of the two knows how much this script might be worth. How can either party determine its reservation price, let alone agree on a mutually acceptable deal?

---

[1] Source: http://www.boxofficemojo.com Worldwide gross box office revenues of the 450 most successful movies in 2008. Less successful movies than *$ 9.99* will possibly not have even made it to this database.

[2] These two examples illustrate the general problem with these models. The best fitting regression model using predictors from the past to explain US revenue has a (non-adjusted) R^2 of only .446. Even for the past, more than half of the variance remains unexplained. The model predicts an average revenue of 2.5 Mio $, but the 95% confidence interval runs from a negative revenue of 89.6 Mio $ to a positive revenue of 94.5 Mio $ (Simonoff and Sparrow 2000:Table 1).

In principle, there is an easy way out. Instead of agreeing on a fixed price, the parties could write a success-contingent contract that gives the artist a defined share of the final profit; in the industry, these are called royalty deals. While such deals are indeed very common with stars, they are less frequent with ordinary artists (Chisholm 1993; De Vany 2004:245). In fact, such a success-contingent arrangement would meet neither side's interests. Producers want to be the exclusive residual claimants and do not want to cut into their managerial freedom. Artists usually have nothing but their human capital for a living and, like ordinary workers, many do not want to bear market risks.

The German legislature explicitly reacted to this situation. It starts from the observation that, in this legal order, "buyout" contracts are widespread. In such contracts, the author sells the right to commercially exploit her creative work in any possible way and using any imaginable technology, against a fixed upfront fee.[3] The German legislature thinks that this situation calls for regulation and stipulates:

> *"If the owner of a copyright has granted a licence such that the fee, in the light of the entire relationship between the parties, is grossly disproportionate with regard to the proceeds from the work, the buyer is obliged to agree, upon the author's request, to a change in the contract such that the seller receives an additional remuneration, reflecting what is her appropriate share under the given circumstances."*[4]

This legal provision seems odd in at least two respects: First, the provision is unbalanced in the sense that it does not grant the buyer the right of an ex-post discount in case the initial price was disproportionately high. Second and more fundamentally, why should the legislature interfere at all in a situation where two parties have voluntarily agreed on a mutually beneficial contract?[5]

---

[3] BT Drs. 14/8058, 1; BT Drs. 14/6433, 9/10/11.

[4] §32a I 1 UrhG[4] (German Copyright Law), our translation. § 32a III 1 UrhG states that the right is not waivable. For an overview of jurisprudence see (Wandtke and Bullinger 2009:§ 32a UrhG).

[5] California has a rule in the same spirit, yet using a different regulatory technique. Under Sec. 986 Cal. Civ. Code, whenever a work of fine art is sold, the artist has a non-waivable statutory right to at least 5 % of the selling price. This provision not only applies if a gallery sells the work, but also if it is sold in an auction or directly by the previous private owner.

In the legislative materials, this provision is introduced as a means to "restore fairness" between the parties[6] in case, given the success on the market, "ex post" the remuneration agreed upon "ex ante" seems inappropriately low.[7] So far, however, there have been no attempts to analyse the effect of the provision empirically. This paper proposes an experimental approach to tackle the following two questions: (1) To what extent does the provision "restore fairness"? (2) How does the provision affect the market for copyrights in terms of (a) number of deals closed and (b) average purchase prices?

Our experimental results show that, upon introducing the provision, more deals are struck, even though offer prices are lower. In addition, the provision leads to a reduction of perceived ex-post unfairness for buyers, but not for sellers. The latter is remarkable given that the experimental umpires, meant to represent the judiciary, apply a fairness norm that clearly benefits sellers.

The paper is structured as follows. Section II presents the design of the experiment and in section III we develop our predictions. In section IV we report the experimental results. Section V concludes.

II. Experimental Design

In stylized fashion our two treatments reflect the interaction on copyright markets *without* (baseline B) and *with* the essence of the German provision (treatment provision P).

A. Baseline (B) treatment

Two roles are randomly assigned to participants: they are either buyers or sellers.[8] This role stays fixed over the entire treatment. The design of the baseline is best illustrated in the flow chart of Fig.1.

[6] BT Drs. 14/8058, 19.
[7] BT Drs. 14/8058, 16.
[8] The bargaining situation is framed neutrally as one of buying and selling a commodity (following Hoffman, McCabe et al. 1994).

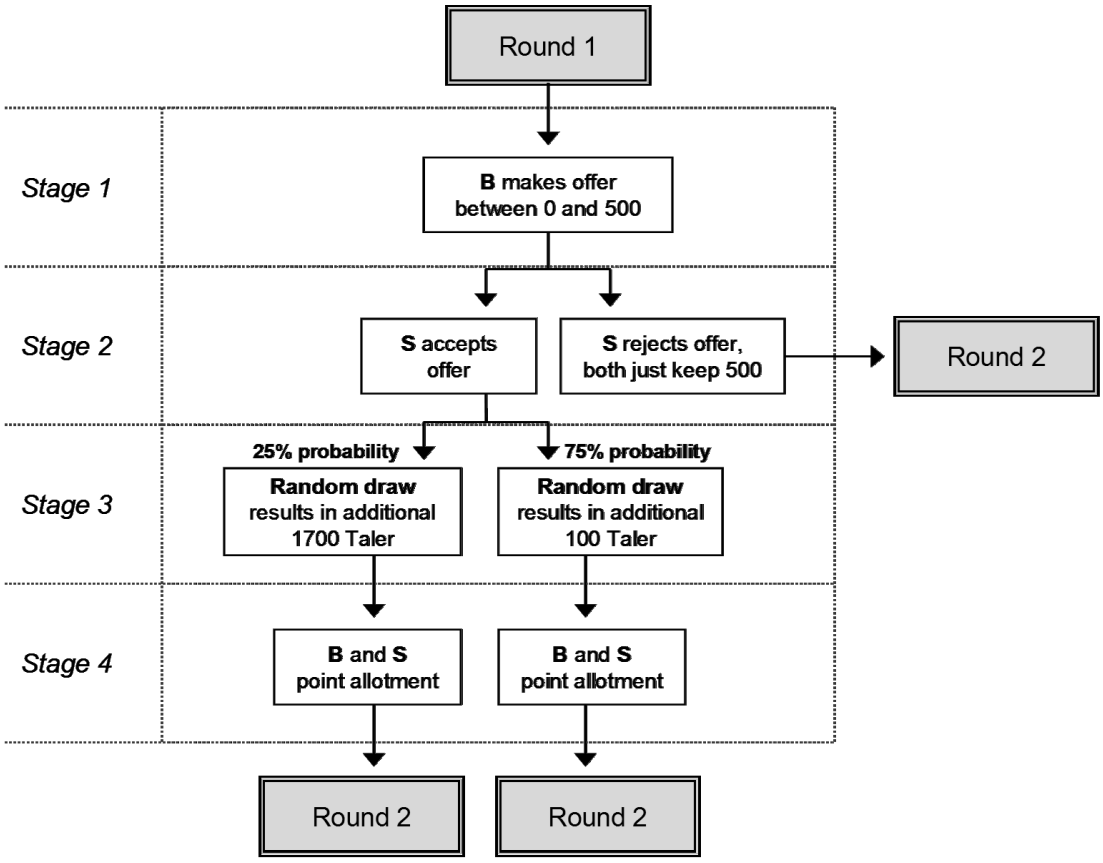


Figure 1: Design of (B) Baseline
Note: The experimental currency is Taler. At the end of the experiment, Taler are converted into Euro at a rate of 1000 Taler = 1 €.

The baseline consists of 8 rounds with a maximum of 4 stages per round: (1) offer stage, (2) acceptance stage, (3) random draw, (4) punishment stage. After each round, buyers and sellers are rematched. We implement a perfect stranger protocol, i.e. a given seller never meets the same buyer twice and vice-versa. Both the matching protocol and the number of rounds are common knowledge to all participants.

At the beginning of each round, both players receive an endowment of m=500 Taler, which is the currency used in the experiment. In stage 1, the buyer has the opportunity to make an offer p in order to purchase commodity C. At this point,

neither the seller nor the buyer know the true value v of C. They only know that with probability $\lambda = .25$ the commodity is worth v_{hi}=1700 Taler, while with the complementary probability $1 - \lambda = .75$ it is only worth v_{lo}=100 Taler. As Figure 2 illustrates, in a stylised fashion this lottery reflects the distribution of earnings in the movie industry.

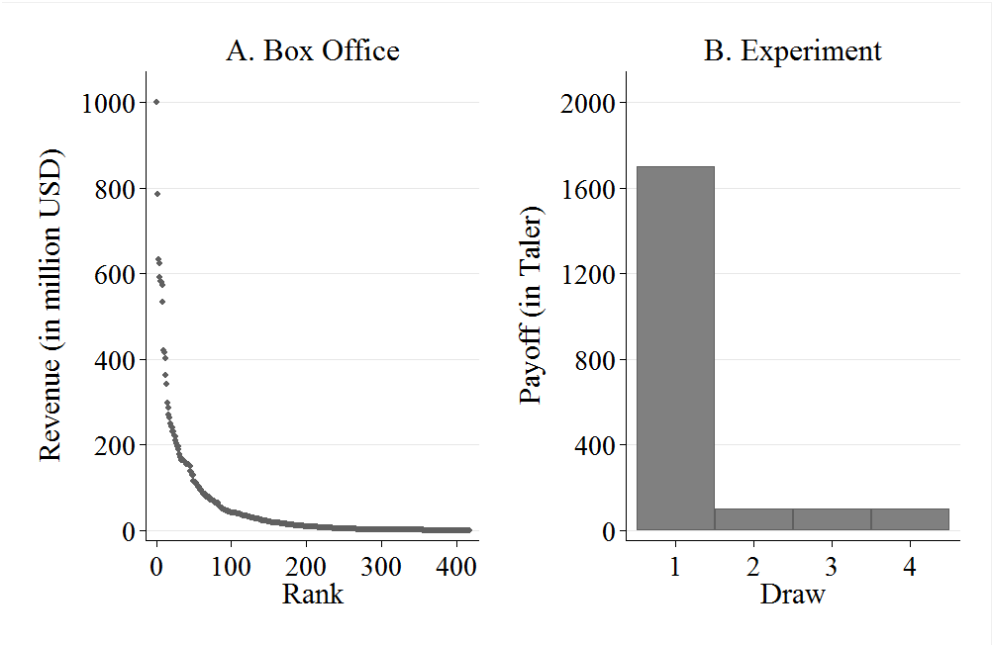


Figure 2: Distribution of earnings

Note: Panel A depicts the worldwide gross box office revenues of the 450 most successful movies in 2008, according to boxofficemojo.com. Panel B illustrates the distribution of earnings in the experiment: In 1 out of 4 cases the commodity is worth 1700 Taler, and in 3 out of 4 cases the commodity is worth 100 Taler.

If the seller rejects the offer p in stage 2, the round is over. In this case there are no gains from trade. Both players just keep their respective endowments of 500 Taler. If the seller accepts, the offer price is immediately transferred from buyer to seller. In stage 3, a random device determines the value of the commodity. In stage 4, both players learn the outcome of the random draw that determines whether the contract is a financial success and have the authority to reduce their counterpart's earnings. The punishment technology is linear (following Fehr and Gächter 2002), and the fine-to-fee ratio is 3, i.e. one Taler spent by the buyer (seller) on punishment reduces the

seller's (buyer's) income by 3 Talers. After stage 4, the round is over and the next round starts. In the instructions, punishment is neutrally labelled "point allotment".

In brief, the parties play an ultimatum game, with the twist that the true value of the commodity is unknown to both the proposer (buyer) and responder (seller). Ultimatum bargaining captures the essence of the situation how the German legislature perceived it, and to which it intended to react by the "bestseller paragraph": the vast majority of copyright owners are offered a fixed license fee on a "take-it-or-leave-it" basis[9].

In the ultimatum game, if the responder rejects an offer, this can be interpreted as punishment (Güth 1995). This might explain why, to the best of our knowledge, no ultimatum game with an explicit punishment stage has been tested. Yet to understand the effect of the provision, it is crucial to have a measure for *ex-post* feelings of (un)fairness (i.e. after the actual value of the commodity has been revealed), in addition to the *ex-ante* indicator of fairness concerns, i.e. rejection of the offer. After all, the German legislature believes that authors accept unfavourable deals too easily as long as they have no clue of the likely value of their works, but feel treated unfairly if their work, after it has been marketed, turns out a success.

To measure ex post fairness attitudes, we capitalise on a tool that is standard in the experimental public goods literature, namely costly, simultaneous punishment. In that strand of the literature punishment is motivated as a technology to discipline freeriders and thereby induce higher levels of cooperation (Fehr and Gächter 2000). However, punishment chiefly is a technology for fairness-driven retaliation (Falk, Fehr et al. 2005). People are willing to incur a cost in order to express their discontent with somebody else's behaviour. Therefore punishment is a good proxy for hurt feelings of fairness (for a similar approach see Fehr, Hart et al. 2010). This interpretation is particularly plausible in our context. The perfect stranger protocol excludes that a participant will herself derive a pecuniary benefit from disciplining a free-rider. Strategic punishment is thus not possible in this design.

[9] BT Drs. 14/6433, 8: "Today, in the Federal Republic of Germany many individual contracts concluded between copyright owners and media industry are not based on collectively negotiated tariffs; copyright owners have to accept standard form contracts prepared by industry on a take it or leave it basis" (our translation).

An alternative to punishment would be a non-incentivized survey instrument, meant to elicit sellers' and buyers' discontent. For the following reasons we prefer the present design. First, any non-incentivized measure is "cheap talk" and therefore potentially unreliable. Since otherwise our experiment is set up in the tradition of experimental economics, we want to be in line with that tradition. Second psychologists, who routinely work with survey instruments, would not trust a mere set of straightforward questions. They instead would want to see standard validation tests, like test retest reliability, or the performance of the test in different environments. There are established psychological measures of related constructs, like the measure for the affective evaluation of risky outcomes by Mellers et al. (Mellers, Schwartz et al. 1997; Mellers, Schwartz et al. 1999). Yet these measures were developed for one-person gambles, while in our experiment participants play a strategic game. Given the rich literature on social preferences (see, for example, Fehr and Schmidt 1999; Bolton and Ockenfels 2000), it seems quite likely that feelings of discontent differ if a poor outcome for this player is associated with a good outcome for another player, and is even caused by this other player's behaviour.

B. Provision (P) treatment

The German provision introduces a third actor, besides the buyer and the seller of the copyright, the court. The court has jurisdiction to intervene, upon the request of the author, in the unlikely but possible event of the work being a big success in the market. The statute does not give the court any guidance about how to change the contract. It thus introduces a discretionary element, with the ensuing ex post uncertainty. In our experimental design we attempt to capture this element of the German solution by introducing a third player: the umpire. As in the case of real judges and juries, the experimental umpire's earnings are unrelated[10] to the decisions she takes; she is paid a fixed fee. This is known to all players. As with the courts, an applicant does not in advance know with certainty how the umpire will decide. The applicant only knows the abstract decision rule. We ask the umpire to determine the "appropriate purchase price" , exactly as in the statutory provision. For the sequence of stages, the reader is referred to the flow chart of Fig.3.

[10] In Germany, judges are appointed at the beginning of their careers and then enjoy the status of civil servants with lifetime tenure. This arrangement is meant to shield them from being held personally liable for the decisions they make.

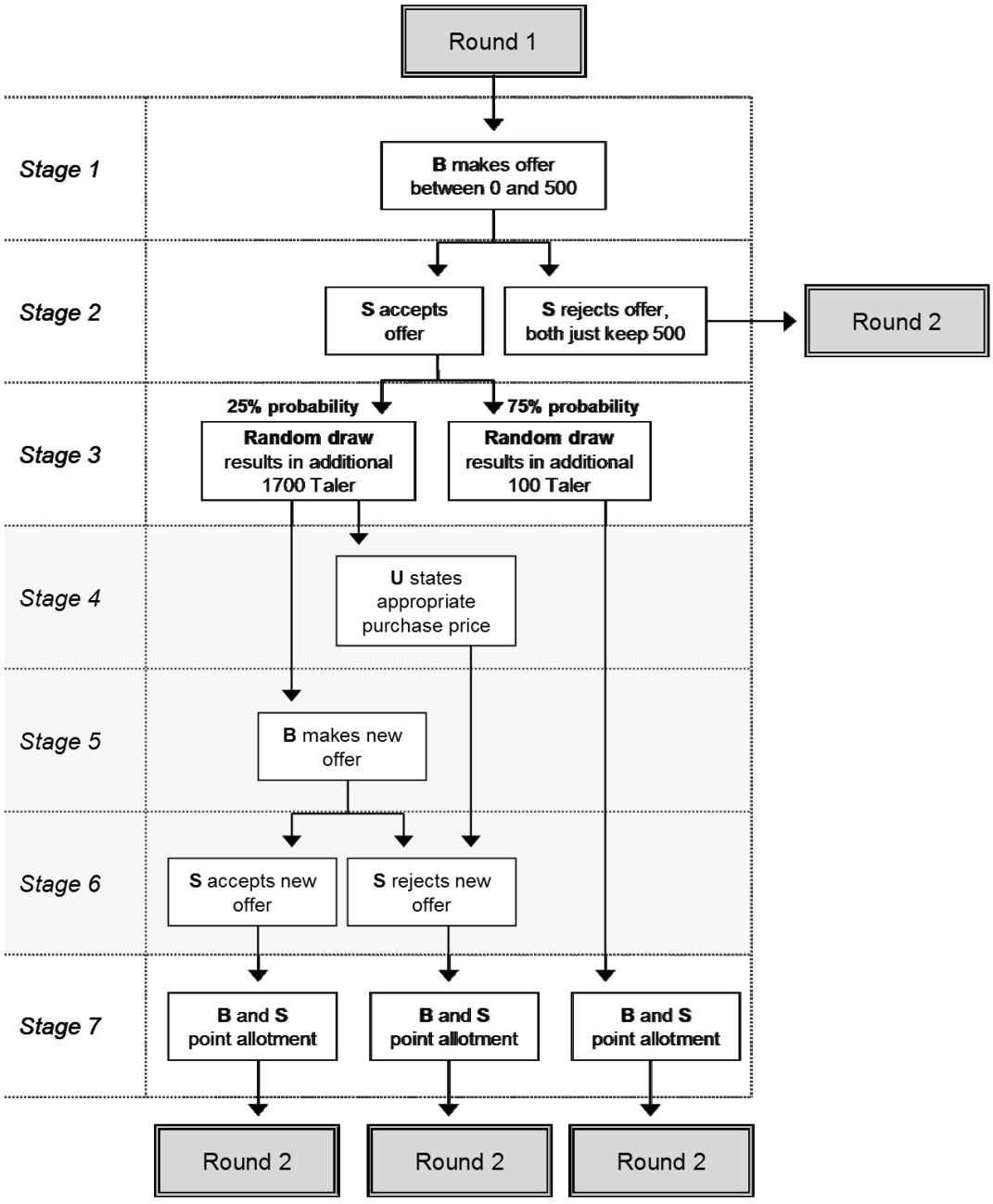


Figure 3: Design of Provision (P) Treatment

Stages 1, 2, and 3 of the P-treatment are identical to the baseline. If Nature draws v_{lo}=100, stage 4 is the punishment stage, so that in fact nothing changes compared with the baseline. If the random draw has determined the value of the commodity to be 1700, there are three additional stages, representing the main features of the provision from German copyright law. If v_{hi}=1700, the punishment stage is deferred to stage 7. In stage 4, the umpire is informed about the purchase price and the random draw and asked to determine an "appropriate purchase price" p_u for the commodity.[11] She may choose any price between 0 and 1700 Taler. Her own payoff does not depend on her choice. Moreover, p_u is not revealed to buyer and seller until stage 7.

Stages 5 and 6 repeat the negotiation protocol from stages 1 and 2. The buyer can make a new offer p_2, which the seller is free to accept or to reject. If p_2 is accepted, it replaces p. If p_2 is rejected, the umpire's "appropriate purchase price" p_u becomes effective and replaces p. Buyer and seller learn about p_u if and only if renegotiation fails. Finally, stage 7 gives both players an opportunity to simultaneously punish each other, just as stage 4 in treatment B.

In reality, the court intervenes only if renegotiation fails. Yet conditioning the decision of the umpire on the rejection of the second offer would have severely reduced the number of data points. This is why we ask the umpire to decide already in stage 4, but we keep her decision confidential until renegotiation fails. We are aware of the fact that this manipulation makes the risk of losing in court more salient. Other experiments have demonstrated that subjects attach more weight to an event if they know that they are betting on the past, rather than betting on the future (Rothbart and Snyder 1970; Ladouceur and Mayrand 1987; Brun and Teigen 1990; Heath and Tversky 1991). However, in our setting both players face the same uncertainty. If the buyer offers too little in stage 5, she risks losing much more in court. Likewise, if the seller rejects a good offer, she risks getting much less in court. Consequently, the stage 4 manipulation might make successful renegotiation somewhat more likely, but it is unlikely to bias the renegotiation outcome if a deal is struck.

[11] The German original of the instructions reads. "Spieler S entscheidet, welchen Kaufpreis er angemessen findet".

C. Procedures

The experiment was run in April 2009 at the Laboratory for Experimental Economics of the University of Bonn using the experimental software z-Tree (Fischbacher 2007). Subjects were randomly invited with ORSEE (Greiner 2004) from a subject pool of approximately 3500 participants. 48 student subjects from different majors participated[12], 27% of which were female. To allow for within-subject comparisons, each subject played first the baseline and then treatment P.[13] Before every treatment participants received paper instructions and answered a set of control questions (see Appendix). Sessions lasted about one and a half hours. In addition to the earnings that depended on their performance in the experiment, participants received a show-up fee of 2 €. On average, participants earned 12.26 €, with 5.34 € from the baseline (6.04 € for buyers and 4.66 € for sellers), and 4.91 € from the P-treatment (5.13 € for buyers, 5.60 € for sellers, 4 € - fix - for umpires).[14]

III. Predictions

The aim of this paper is to experimentally test the effect of a specific institutional intervention –the "bestseller paragraph"– in a stylised market for copyrights. Economic theory provides a roadmap for the subsequent analysis of our experimental results.

A. Standard Framework

Under standard assumptions, i.e. common knowledge of the fact that agents are selfish, risk neutral and apply backward induction, the equilibrium solution to the baseline is straightforward: Neither of the parties uses costly punishment in stage 4 (since it has a cost, but no pecuniary benefit) and the seller accepts any positive offer

[12] 12.5% of participants were law students and 25% economics students.

[13] We implement this specific sequence of treatments because we explore whether introducing this provision is desirable. We are not interested in the mirror question: what would happen, were the provision abolished?

[14] One-third of the buyers in baseline B and one-third of the sellers in baseline B became umpires in treatment P. The remaining buyers and sellers stayed in their roles. Buyers, sellers and umpires were re-matched after each round, such that no triad played together more the once and no buyer met the same seller twice. As we show later, umpire behaviour turned out not to be correlated with having previously experienced the role of the buyer or the seller.

in stage 2 (since any positive outcome is better than nothing, and the design of the experiment gives all the bargaining power to the buyer). This is anticipated by the buyer, who therefore offers the smallest positive price $p>0$ in stage 1 (since she is sure it will be accepted).

In the P-treatment, we should again observe no punishment in stage 7 (or stage 4, when the commodity has low value), since for a purely profit maximising agent the benefits from punishing are zero[15]. Whether the seller accepts or rejects the new offer p_2 in stage 6 depends on what she believes will be the umpire's "appropriate purchase price" p_u. The seller will only accept the new offer if $p_2 \geq E(p_u)$. On the other hand, the buyer will not offer more than $E(p_u)$.

In stage 4 the umpire has to state her appropriate purchase price p_u. Since the umpire's decision is not incentivised, we cannot predict her behaviour with rational choice theory. Instead, we assume three stylised types of umpires, each of whom having a different conception of fairness: *Umpire L* has a libertarian mindset and thinks that the initial agreement should simply be kept, hence $p_u^L=p$. *Umpire EA* is an egalitarian who looks at the situation from an ex-ante perspective, i.e. before the veil of uncertainty is lifted. Her definition of a fair price would anchor on the expected value of the lottery, yielding $p_u^{EA}=E(v)/2=\frac{1}{2}((1-\lambda)v_{lo}+\lambda v_{hi})$. *Umpire EP* also holds an egalitarian attitude but rather considers the ex-post situation, i.e. after nature has determined the value of the commodity to be high. Therefore, her appropriate purchase price is $p_u^{EP}=v_{hi}/2$.

With umpire L, the P-treatment becomes identical to the baseline. In stage 2, the seller will accept any positive price and in stage 1, the buyer will just offer the minimum. With umpire EA, the seller accepts an offer p if and only if $0 \leq (1-\lambda)p+\lambda p_u$. Inserting the parameters of the experiment, we obtain $p \geq -E(v)/6=-83.33$ as the minimum acceptable price. Hence, with the prospect of an ex-ante egalitarian umpire even negative offer prices appear acceptable to the seller. This is even more so with umpire EP, where, following the same logic, the minimum acceptable price in stage 2 would be $p \geq -v_{hi}/6=-283.33$. Since we exclude negative prices by design, under standard assumptions we should not observe any difference

[15] Recall that strategic/educative punishment is pointless with perfect stranger matching.

between the two treatments. As we have just shown, this prediction is independent of the expected type of umpire. Hence, the predictions can be summarized as follows:

S1. Same offer prices in P as in B, namely the smallest possible price.
S2. Same acceptance rate in P as in B, namely 100%.
S3. Same punishment behaviour in P as in B, namely zero punishment, for both buyers and sellers.

B. Behavioural Framework

Once we relax standard assumptions and allow for well-documented behavioural regularities, predictions are less clear cut. Several behavioural effects are likely to play a role in our setting, most notably risk aversion, loss aversion and social preferences. So far, however, there is no general theory that integrates all of them. It is beyond the scope of this paper to propose a comprehensive theory of social preferences under risk and uncertainty, which would be needed in order to formally derive behavioural predictions. We can however use the individual behavioural building blocks in order to qualify the predictions derived on the basis of the standard model.

Numerous studies in psychology and behavioural economics have shown that most people do not behave like rational money-maximisers but that they care about fairness (Kahneman, Knetsch et al. 1986; Konow 2000). This behaviour has most prominently been explained by social preferences (Fehr and Schmidt 1999; Bolton and Ockenfels 2000; Charness and Rabin 2002). From many previous experiments with ultimatum games it is known that participants often reject deals that would make them strictly better off in monetary terms (Güth, Schmittberger et al. 1982). According to a survey by (Camerer 2003:49), if the offer does not exceed 20 % of the pie, it is rejected in approximately 50 % of the cases. Anticipating this type of seller behaviour, buyers would have to make more substantial offers if they do not want to miss the opportunity to obtain the gains from trade. In the literature on ultimatum games, median offers were in the order of 40 – 50 % of the pie and mean offers were in the order of 30 – 40 % of the pie (Camerer 2003:49).

Our baseline setting, however, differs from the standard ultimatum game in one important respect: when buyer and seller negotiate, the value of the pie is uncertain to both parties. However, only one party, the buyer, bears the risk. Our design thus

involves two-sided uncertainty and unilateral risk-bearing. To the best of our knowledge, such a paradigm has not been experimentally tested.[16]

In our design, the fact that the commodity trades before its true value is revealed implies that the buyer obtains a risky asset in exchange for a fixed fee paid to the seller. Empirically we know that most people are risk averse (Holt and Laury 2002; Fullenkamp, Tenorio et al. 2003; Dohmen, Falk et al. 2005). Risk aversion would reduce buyers' willingness to pay for a risky commodity compared to a deterministic commodity with identical expected value. Provided $p>v_{lo}$, the buyer even runs the risk of making a real loss. In that event, the effect of loss aversion (Kahneman and Tversky 1979) might even go beyond the effect of mere risk aversion.

In contrast, the seller's decision does not involve risk. Hence, we should not expect risk aversion to affect her willingness to accept a certain offer price; at least not directly. Indirectly, risk aversion might influence sellers' behaviour through social preferences. If the rejection of low offers is driven by fairness concerns, we should expect that sellers are sensitive to the fact that buyers unilaterally bear the risk. This should lower the acceptance threshold of sellers, compared to deterministic ultimatum games. The dampening effect should be all the more pronounced given the distribution of gains is highly skewed. Thus, while we would expect offers in the baseline to be considerably above the minimum price, we also expect prices to be below the typical result from deterministic ultimatum games.

We have shown above, for the case of standard rationality assumptions, that the acceptance threshold of the seller depends on the beliefs subjects hold about the fairness conception of the umpire: the acceptance level did not change if the seller expects an umpire of type L, but decreased for type EA and dropped even further for type EP. For the parameters of the experiment, however, this would not trigger a treatment effect since we exclude negative prices. Yet, if we account for the empirical regularity that, due to social preferences, prices need to be considerably above the minimum in order to be accepted, the predictions for the effect of the provision might indeed change. More specifically, if prices in the baseline are sufficiently high and if subjects anticipate umpires of type EA or EP, the commodity should trade at lower prices in treatment P than in the baseline. For umpires of type L, we would again

[16] There are, however, ultimatum games with one-sided uncertainty. See for instance Mitzkewitz and Nagel (1993) and Abbink, Bolton et al. (2001).

expect no treatment difference. Consequently, unless participants hold extreme beliefs about the distribution of umpire types, we expect lower average prices in the treatment.

The provision's effect on the number of deals closed is less clear. From a normative perspective, this is the most important measure. If copyright does not trade, all potential welfare gains are foregone. Behavioural theory predicts two effects that point in opposite directions. Just as umpires of type EA and EP make low offers more acceptable to sellers, they also reduce buyers' willingness to pay high prices since the v_{hi} prospect becomes less attractive in view of potential ex-post redistribution. Offers above v_{lo} appear even more unattractive since buyers would then suffer a loss in 75% of the cases, never knowing how much the umpire will let them keep in case of v_{hi}. If the decline in the sellers' acceptance threshold outweighs the decrease of buyers' willingness to pay, we should observe more deals being closed in the P-treatment than in the B-treatment.

Equally open is the effect of the provision on ex-post unfairness. By ex-post unfairness we mean the fairness sentiments after the true value of the commodity has been revealed, which we attempt to measure by costly, simultaneous, non-strategic punishment. In the baseline this takes place in round 4; in the P-treatment it is elicited in round 4 if $v=v_{lo}$ and in round 7 if $v=v_{hi}$, hence after renegotiation and after the decision of the umpire has been revealed. As stated above, the main expected effect of the provision is to reallocate profit from the buyer to the seller in case umpires are of type EA or EP[17]. In addition, the anticipation of those types of umpires might lower the acceptance threshold of sellers and consequently average selling prices. In that case buyers would be better off with the provision if $v=v_{lo}$ (due to lower prices) and worse off if $v=v_{hi}$ (due to ex-post reallocation), and vice-versa for sellers. Just as the provision might simply reallocate profits but not affect the number of deals closed, it might also merely reallocate ex-post discontent. If lower profits lead to more discontent, we should observe buyers (sellers) punishing less (more) in the P-treatment than in the baseline if $v=v_{lo}$ and more (less) if $v=v_{hi}$. However, for the provision to not only reallocate perceived unfairness but to actually "restore fairness" as the legal literature claims, overall punishment would have to go down. In sum, we thus have the following predictions:

[17] As we have also mentioned above, the provision should have no effect on ex-post unfairness if umpires are of type L and this is correctly anticipated by the subjects.

B1. Lower offer prices in P than in B.
B2. Higher acceptance rate in P than in B indicates that the provision has a stronger effect on reducing sellers' acceptance threshold than on decreasing buyers' offer prices.
B3. Less buyer punishment in P than in B for $v=v_{lo}$ and more for $v=v_{hi}$; and vice-versa for sellers.

IV. Results

In the following we discuss the impact of the provision in terms of ex-ante prices and number of deals closed as well as ex-post renegotiation, umpire decisions and expression of discontent according to our experimental results. Table 1 gives an overview of the data collected in the experiment.

Table 1: Data Structure of the Experiment

	Subjects	Buyers	Sellers	Umpires	Rounds	Offers	Deals	v_{lo}	v_{hi}
Baseline	48	24 (16)	24 (16)	0	8	192 (128)	141 (98)	105 (74)	36 (24)
Provision	48	16	16	16	8	128	106	82	24

Note: The numbers in parentheses denote the number of observations when we restrict the sample to those buyers and sellers who stayed in the same role during baseline and provision treatment. In all regressions we must work with the restricted sample. For the descriptives we use all data. Means and standard deviations look almost identical if we only use the restricted sample.

A. Prices and Deals

Fig. 4 Panel A shows that mean offer prices in stage 1 are way above the price of 1 taler predicted by money maximisation. More importantly, both mean offer prices and mean accepted prices are considerably higher in the baseline than in the P-treatment.

These differences are significant.[18] Note that in both cases average offer prices are above 100, namely 129 in the baseline and 104 in the P-treatment. Hence, in both cases the average buyer was willing to run the risk of incurring a loss in the v_{lo}=100 case, in the interest of reaping high profits if v_{hi}=1700. Just as we hypothesised, the buyers' willingness to pay high initial prices was reduced by the introduction of the provision. Moreover, offered and accepted prices are closer in the provision treatment, which hints at a more efficient functioning of the market. Indeed, the probability of acceptance in stage 2 rises from 73% in the baseline to 83% in the provision treatment. This suggests that the provision managed to reduce the acceptance threshold of sellers relatively more than the willingness to pay of buyers. As a result, more deals were closed in the presence of the provision than in the baseline.

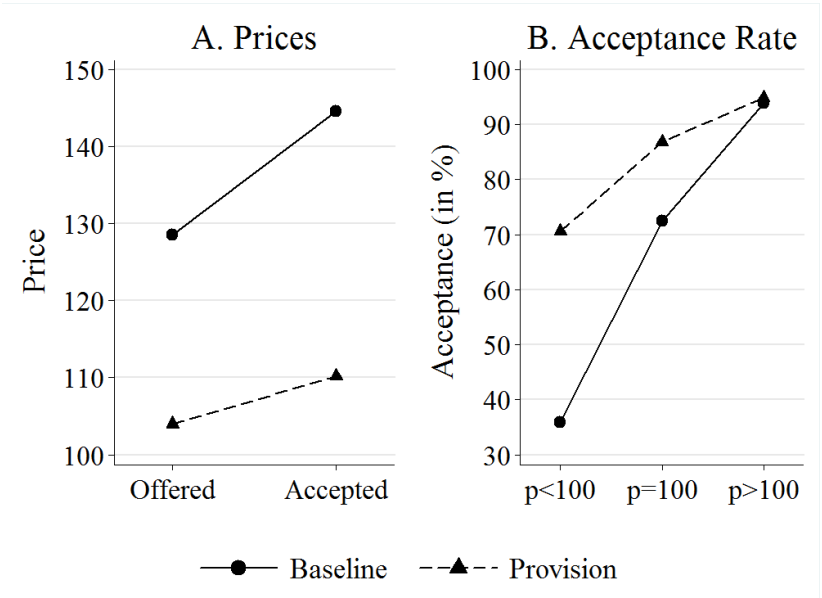


Figure 4: Deals reached in both treatments

Note: Panel A illustrates the average prices offered in stage 1 and accepted in stage 2 in the Baseline as well as in the Provision treatment. In the Baseline (Provision), 141 (106) out of 192 (128) offers were accepted. Panel B distinguishes the acceptance rate depending on the offer price being below, at, or above 100 Taler. The number of observations in the Baseline (Provision) is 53 (51) for p<100, 40 (38) for p=100, and 99 (39) for p>100.

[18] Wilcoxon signed-rank test over mean offer per buyer, using only data from those 16 participants who were buyers in both treatments, two-sided, p = .0525; same for mean accepted offer, p = .0247.

The main reason why more deals were struck is that the provision made low offers more acceptable to sellers. In Panel B we distinguish between offers below, at and above 100. This threshold is important because in the probable event of v=100 buyers are just equally well off with or without the deal if they paid p=100. At p<100 they are sure to make a profit whereas at p>100 they make a loss when v=100. In contrast, depending on the size of p sellers may make larger or smaller profits, yet they never make a loss. The Figure illustrates that while in the baseline buyers had to incur the risk of making a loss in order to strike a deal, the acceptance of risk-free offers nearly doubles in the presence of the provision.

Result 1: The provision decreases the average price paid for the commodity
Result 2: The provision increases the total number of deals reached by making low prices more acceptable to sellers.

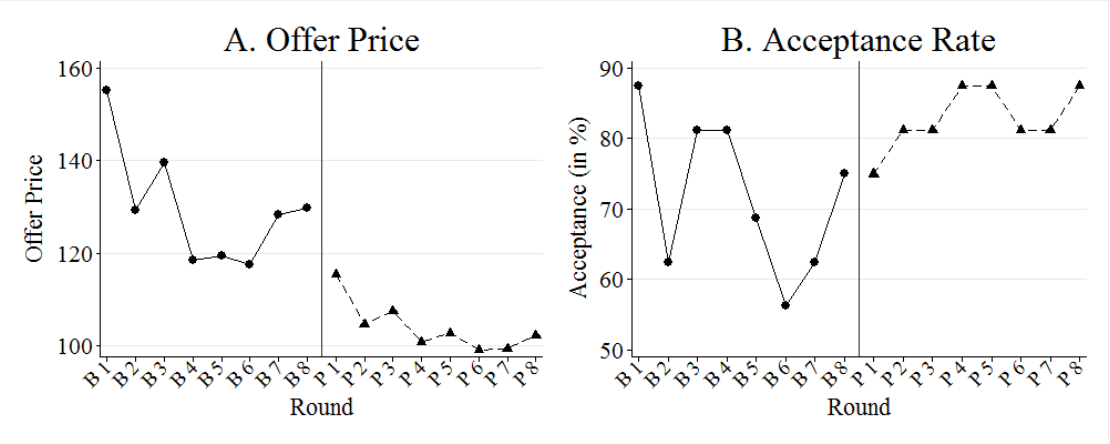


Figure 5: Buyers' Offers and Sellers' Acceptance over Time
Note: The vertical line indicates the change of the institutional regime from Baseline (B) to Provision (P). As we are checking for learning, we only use data from those buyers (panel A) and sellers (panel B) who have been in this role in both treatments.

In principle, the treatment effect could result from learning, rather than from the change in institutions. Specifically lower prices in P could be due to a general downward trend; and the lower acceptance rate could be the consequence of an upward trend. In contrast, if the institutional setting matters, we should observe a pattern break upon introducing the provision. Panel A of Fig. 5 suggests that buyers do indeed learn to make lower offers. Yet learning takes place *within* both treatments,

and the effect of learning does not appear linear. While in both treatments offers are highest in the beginning, they are lowest at some intermediate point, to go up again by the end of each treatment. More importantly, there is a visible level effect when moving from the baseline to the provision treatment. The sellers' acceptance rate of offers, depicted in Panel B, is very volatile over time. There does not seem to be a general time trend.

Table 2 – Treatment Effect on Buyers' Offers Controlling for Learning
Dependent variable: Buyers' Offers

	(1)	(2)	(3)
Treatment (0=B, 1=P)	-25.766***	-25.766***	-25.766***
Time Trend		-2.234^{+}	-12.563*
Time Trend Squared			1.138^{+}
constant	129.734***	140.191***	157.256***
N	256	256	256
Chi Squared	19.66	23.25	26.76

Note: Panel Regression with a random intercept for individuals. $^{+}$ denotes significance level of 10 percent, * denotes significance level of 5 percent, ** of 1 percent and *** of 0.1 percent. The Hausman test is insignificant for all models, hence the Random Effects Model is consistent.

We assess the significance of our results with parametric tests. For the buyers' offers we estimate random effects models that account for subject heterogeneity and report the results in Table 2. As one should expect seeing Fig. 5, in both treatments there is a significant negative time trend. Offers become smaller over time. If we capture the upward movement by the end of the treatment with the square of the time trend (model 3), the negative time trend is significant at conventional levels. Yet adding these controls neither affects the treatment coefficients, nor significance levels. Actually, the treatment coefficients are identical to the third decimal place. Our models suggest that there is indeed learning. Yet learning is clearly independent of the treatment effect.

Table 3: Treatment Effect on the Sellers' Acceptance Controlling for Learning
Dependent variable: Sellers' Acceptance

	(1)	(2)	(3)
Treatment (0=B, 1=P)	1.062*	1.065*	1.065**
Time Trend		-.051	-.153
Time Trend Squared			.011
constant	1.509*	1.740*	1.914^{+}
N	256	256	256
Chi Squared	6.71	7.02	7.08
Log Likelihood	-100.746	-100.570	-100.535

Note: Panel Logit Regression with a random intercept for individuals. * denotes significance level of 5 percent, ** of 1 percent and *** of 0.1 percent. The Hausman test is insignificant for all models, hence the Random Effects Model is consistent.

As the seller's decision to accept or reject the buyer's offer is binary, we report panel logit models in Table 3. Also in this case, the treatment effect is not affected by including a linear or a quadratic time trend. The trend coefficients are insignificant. We have thus no indication of learning.

B. Renegotiation and Umpire Decisions

The very fact that we observe lower prices in the provision treatment suggests that subjects were anticipating some sort of ex-post redistribution towards the seller in case that $v=v_{hi}$, be that through renegotiation or through the umpire's decision. Yet when it comes to splitting the large gain, there is a pronounced self-serving bias. The histogram of second offers in Panel A of Fig. 6 shows that buyers believe they are justified in keeping most of the large gain, while sellers believe they have a right to a large portion. 54% of second offers are rejected. Sellers do not accept any second offers below or equal to 250 and accept only 15% of all second offers below or equal to 500. Seemingly, for second offers to be acceptable, they must be above 500; 82% of such offers are accepted.

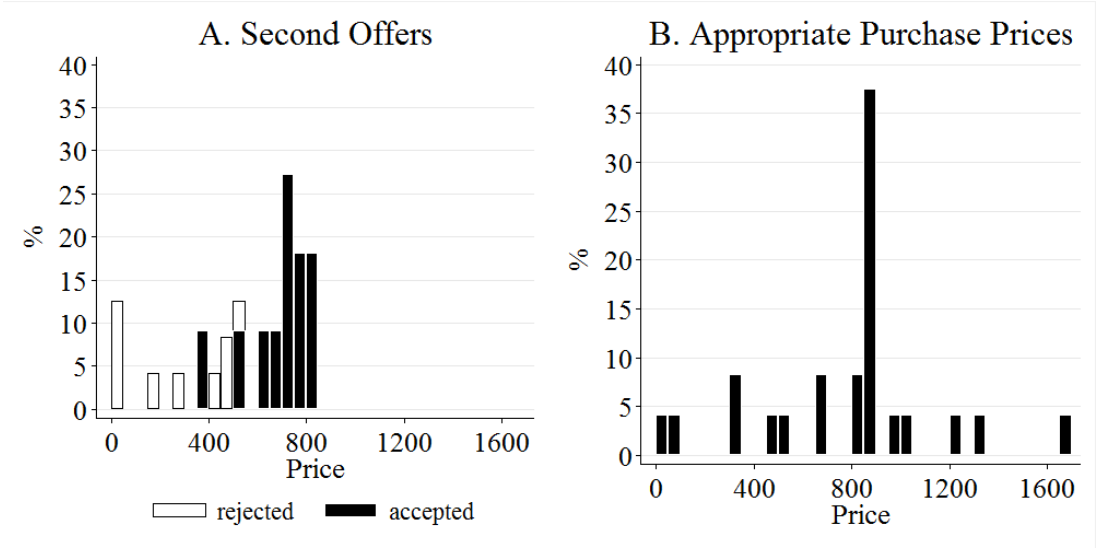


Figure 6: Renegotiation and Umpire Decisions

Note: Panel A shows the second offers made by the buyers in the Provision treatment in case the commodity turned out to have value 1700. The light (dark) bars denote the second offers rejected (accepted) by the sellers. Panel B displays the umpires' choices of "appropriate purchase prices" in case the commodity had value 1700. The number of observations is 24.

Apparently, sellers had a better intuition of how umpires would view the "appropriate purchase price". On average, umpires decided that 767 Taler should go to the seller, which leaves 933 Taler to the buyer. From an ex-post perspective this average umpire decision may appear to slightly favour buyers. However, from an ex-ante perspective it clearly favours sellers: Considering that initial offer prices averaged 104 Taler, buyers could expect to make 0.75*(-4)+0.25*933=230 Taler whereas sellers would receive 270 Taler.[19] The fact that buyers ran the risk of making a real loss does not seem to enter the fairness considerations of the average umpire. If any, the risk premium is negative.

This result is highly surprising. The experiment was designed such that umpires had the same information as buyers and sellers. They knew that the initial offer price was paid in exchange for a lottery ticket which could be worth either 100 or 1700 Taler. They further knew that the game was repeated 8 times so that there were repeated

[19] With average accepted prices the difference is even larger: 225 to 275.

opportunities to land the jackpot. We therefore expected many umpires to focus on the expected value of 500 and judge the fair price to be somewhere around 250, which would be in line with ex-ante equality. We also expected some umpires to confirm the initial offer. Yet, we find no peak whatsoever around 250 and there is only a single instance in which an initial offer was confirmed. Instead, as can be seen in Panel B of Fig. 6, 38% of the umpires behaved like prototypical ex-post egalitarians and precisely split the ex post gain equally between the buyer and the seller. Moreover, our within design implied that half of the umpires were former buyers and the other half former sellers. However, previous experience as buyer or seller had no significant effect on the umpire's fairness preferences.[20]

Result 3: Ex post, buyers and sellers hold self-serving views about the equitable division of gains from trade.

Result 4: Umpires' choices are predominantly guided by ex-post equality.

C. Ex-Post Discontent

We have seen that the provision increases the acceptability of low offers and thus enhances the incidence of trade. Yet despite the fact that the parties bargain under the shadow of the umpire's decision (cf. Mnookin and Kornhauser 1979), more often than not renegotiation fails. This already suggests that there is a potential conflict which could support legal intervention. The German legislature claimed that the provision was necessary to "restore fairness".[21] We measure the seriousness of perceived unfairness by the amount of experimental money the two parties are willing to burn in order to express their discontent, and inflict harm on the other party.

Buyers punish more severely in the baseline: 10 Taler per buyer per round versus 7 Taler in the provision treatment.[22] They also use the punishment option more frequently in the baseline (21% of all possible cases) than in the provision treatment (12%). Descriptively, this effect appears to be driven by the fact that buyers punish with higher probability (25%) if they have made an offer above 100, i.e. if they have

[20] Descriptively, former buyers on average even gave more to the seller (873 Taler) than former sellers (691 Taler).

[21] BT Drs. 14/8058, 19.

[22] Wilcoxon signed rank test over mean punishment in the baseline and the provision treatments, for those 16 participants who were buyers in both treatments, $p = .0037$.

accepted the risk of a real loss. With offers at or below 100, the punishment probability is only 8%. As we saw in Fig. 2, deals at or below 100 are much more probable in the provision treatment than in the baseline. In addition, this difference is considerably more pronounced in the provision treatment (3% at or below 100 and 30% above 100) than in the baseline (15% to 24%). This shows an interesting relationship between the ex-ante and the ex-post dimension of fairness: In the baseline, offers below 100 appear (ex-ante) unfair to the sellers and offers above 100 seem (ex-post) unfair to the buyers. The provision increases the acceptability of low offers for the sellers. This is anticipated by the buyers who make more low offers, which in turn reduces their ex-post discontent. Table 4 underpins the robustness of the provision's effect on buyers' use of punishment. All models show a large and significant treatment effect.

Table 4: Explaining Buyer Punishment

Dependent variable: Use of punishment by the buyer (0=No, 1=Yes)

	(1)	(2)	(3)	(4)	(5)
Treatment (0=B, 1=P)	–1.634*	–1.638*	–1.666*	–1.948*	–5.514*
Price offered		0.007			
Value = 1700 (0=No, 1=Yes)		1.350	1.294	0.895	1.130
Offer > 100 (0=No, 1=Yes)			0.313	0.452	–1.355
Treatment x Value=1700				0.952	-1948
Treatment x Offer>100					4.392
constant	-3.517**	-5.243**	-4.193**	-4.122**	-3.649**
N	204	204	204	204	204
Chi Squared	5.334	7.541	6.989	6.695	6.519
Log likelihood	–47.435	–45.523	–46.055	–45.878	–43.758

Note: Panel Logit Regression with a random intercept for individuals. * denotes significance level of 5 percent, ** of 1 percent and *** of 0.1 percent. The Hausman test is insignificant for all models, hence the Random Effects Model is consistent.

Sellers make very little use of punishment. Only in 16 cases, i.e. only in 6% of all possible instances, did sellers punish buyers. In the baseline, they punished twice when the commodity had value 1700, in the provision treatment they punished in a single case. Patently, sellers' fairness sentiments are not offended. For them, it does

not cause a problem if, eventually, the commodity has high value, although the chance of making so much money was not reflected in initial offers. So, apparently, sellers do not feel ex post discontent in the first place and, as a consequence, there is no margin for the provision to improve upon the baseline situation.

Result 5: The provision reduces buyers' ex-post discontent, but does not affect sellers' feelings of perceived ex-post fairness.

V. Conclusion

In this experiment we have compared two institutional arrangements for the market of copyrights. The two most important features of this market are the skewed distribution of earnings and their unpredictability. In the first institutional setting, the copyright must be traded under the veil of uncertainty for a fixed fee. In the unlikely, but possible event of high success in the market, the licence fee is nonetheless binding for the two parties. This situation reflects a type of contract that is very common in the market for copyrights, most notably between large production firms and little known copyright owners. In contrast, the second setting introduces renegotiation in the shadow of legal intervention. In case the work turns out a bestseller, the artist may appeal to a third party who is entitled to adjust the fee. This situation corresponds to a provision from German copyright law, the so-called "bestseller paragraph".

In a market characterised by high uncertainty about the value of the traded goods, conflicting fairness norms between buyers and sellers are amongst the biggest obstacles to trade. Yet even when the parties have reached an agreement in the first place, substantial discontent may arise as soon as the true value of the commodity is revealed. In this experiment we have measured fairness ex ante by looking at the acceptance of initial offers, and fairness ex post by analysing the expression of discontent through punishment. In addition, we had a third party judging fairness, the so-called umpire, who was free to choose among or compromise between competing fairness norms.

Our first finding is that, in the presence of the provision, copyrights trade at lower prices. Second, they trade more often as the acceptance level of sellers is reduced more than the willingness to pay of buyers. Higher acceptance is a strong indicator of enhanced ex-ante fairness. Interestingly, so far the legal discourse has been totally

neglecting the provision's effect on the market outcome. In contrast, both the German legislature and the legal literature have concentrated on the ex-post dimension of fairness, claiming that the main function of the provision is to "restore fairness" between the two parties. Our results suggest that, indeed, the provision reduces perceived unfairness for buyers. Rather surprisingly, though, we do not find a similar effect for sellers. This might of course be due to the fact that, in our design, sellers are not personally attached to the commodity they trade. Copyright combines a property right with a moral right. The latter is absent from our design.

In the provision treatment, a third party, the so-called umpire, was asked to determine her "appropriate purchase price" in case the commodity had a high value. Even though our design was rather prone to highlight ex-ante equality (by repeating the game 8 times, by having umpires experience the roles of buyer and seller, and by telling subjects the exact probabilities so that they could calculate expected values), ex-post equality turns out the umpires' single distinct fairness norm. If probabilities were not known, as is the case in reality, ex-ante equality would possibly be even less appealing. Similarly, there were no umpires whatsoever following a libertarian approach according to which voluntarily closed contracts should simply be kept.

We have used the movie industry as our primary example because we have precise data on this market. Yet, given the neutral frame of our experimental design, the insights from this study should also be relevant to other copyright markets, characterised by comparable unpredictability of earnings, like exhibitions (Skinner 2006) and music (Davies 2002). Other markets with highly skewed earnings, like venture capital, might also be affected by a similar fairness problem.

Obviously, our stylised experimental setting had to abstract from features of reality. For instance to make a movie, a large number of holders of intellectual property rights must contribute. This feature of the market makes ex ante deals with all holders of such rights paramount. One might argue that there is less reason for legal intervention if the number of copyright holders is much smaller (as sometimes with music) or if there is just a single copyright (as sometimes with literature).[23] Yet our data suggest that the provision has effects even if the relationship is strictly bilateral.

[23] In the special case of the music business, many authors are at the same time performers of their work. As a consequence they are often willing to accept very low or even negative license fees, just to get access to a larger audience, which increases concert revenues.

In our experiment, sellers have a single opportunity to sell a copyright to a specific buyer. The buyer has all the negotiation power. In the field, there may be more than one potential distributor, and sellers may have a chance to wait until they come into contact with such a firm. Experimental evidence suggests that in ultimatum games responder competition improves outcomes for proposers, i.e. lower prices, just as bargaining theory would have suggested (Grosskopf 2003). This may well also hold in the media industry. Yet the German provision is by no means contingent on market structure or bargaining power. Our experiment indicates that, at least in this static and bilateral setting the provision is beneficial for sellers, buyers and society. Future work might want to introduce different allocations of market power and test whether under such less favourable conditions the welfare enhancing effect of the provision is substantially reduced. In addition, one could make the sellers' production of the copyright endogenous, to study the dynamic effects of different institutional arrangements. Another extension could allow the buyers' effort to influence the probability of success. One might also want to study to which degree reputation is a substitute for legal intervention.

The German solution discussed in this paper neglects more sophisticated schemes that have been proposed in the economics literature to guide contracting about copyright when success is uncertain (Watt 2006 surveys this literature). Further research might seek to experimentally compare the performance of those alternative schemes. This paper's results highlight the need for a general theoretical framework that integrates social preferences – both of interested and neutral parties – with decision making under risk and uncertainty.

References

ABBINK, KLAUS, GARY E. BOLTON, ABDOLKARIM SADRIEH and FANG-FANG TANG (2001). "Adaptive Learning versus Punishment in Ultimatum Bargaining." Games and Economic Behavior **37**: 1-25.

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

BRUN, WIBECKE and KARL H. TEIGEN (1990). "Prediction and Postdiction Preferences in Guessing." Journal of Behavioral Decision Making **3**: 17-28.

CAMERER, COLIN F. (2003). Behavioral Game Theory. Experiments in Strategic Interaction. New York, NY, Sage u.a.

CHARNESS, GARY and MATTHEW RABIN (2002). "Understanding Social Preferences with Simple Tests." Quarterly Journal of Economics **117**: 817-869.

CHISHOLM, DARLENE C. (1993). "Asset Specificity and Long-Term Contracts. The Case of the Motion-Picture Industry." Eastern Economic Journal **19**: 143-155.

COLLINS, ALAN, CHRIS HAND and MARTIN C. SNELL (2002). "What Makes a Blockbuster? Economic Analysis of Film Success in the United Kingdom." Managerial and Decision Economics **23**: 343-354.

DAVIES, J.A. (2002). "The Individual Success of Musicians, Like that of Physicists, Follows a Stretched Exponential Distribution." European Physical Journal B **27**: 445-447.

DE VANY, ARTHUR S. (2004). Hollywood Economics. How Extreme Uncertainty Shapes the Film Industry. London ; New York, Routledge.

DE VANY, ARTHUR S. and W. DAVID WALLS (1996). "Bose-Einstein Dynamics and Adaptive Contracting in the Motion Picture Industry." Economic Journal **106**: 1493-1514.

DE VANY, ARTHUR S. and W. DAVID WALLS (1999). "Uncertainty in the Movie Industry. Does Star Power Reduce the Terror of the Box Office?" Journal of Cultural Economics **23**: 285-318.

DOHMEN, THOMAS J., ARMIN FALK, DAVID HUFFMAN, JÜRGEN SCHUPP, UWE SUNDE and GERT G. WAGNER (2005). "Individual Risk Attitudes. New Evidence from a Large, Representative, Experimentally-validated Survey."

FALK, ARMIN, ERNST FEHR and URS FISCHBACHER (2005). "Driving Forces Behind Informal Sanctions." Econometrica **73**(6): 2017-2030.

FEHR, ERNST and SIMON GÄCHTER (2000). "Cooperation and Punishment in Public Goods Experiments." American Economic Review **90**: 980-994.

FEHR, ERNST and SIMON GÄCHTER (2002). "Altruistic Punishment in Humans." Nature **415**: 137-140.

FEHR, ERNST, OLIVER HART and CHRISTIAN ZEHNDER (2010). "Contracts as Reference Points. Experimental Evidence." American Economic Review ***: ***.

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.

FULLENKAMP, CONNEL, RAFAEL TENORIO and ROBERT BATTALIO (2003). "Assessing Individual Risk Attitudes Using Field Data From Lottery Games." Review of Economics and Statistics **85**: 218-226.

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

GROSSKOPF, BRIT (2003). "Reinforcement and directional learning in the ultimatum game with responder competition." Experimental Economics **6**(2): 141-158.

GÜTH, WERNER (1995). "On Ultimatum Bargaining - A Personal Review." Journal of Economic Behavior and Organization **27**: 329-344.

GÜTH, WERNER, ROLF SCHMITTBERGER and BERND SCHWARZE (1982). "An Experimental Analysis of Ultimatum Bargaining." Journal of Economic Behavior and Organization **3**: 367-388.

HEATH, CHIP and AMOS TVERSKY (1991). "Preference and Belief. Ambiguity and Competence in Choice under Uncertainty." Journal of Risk and Uncertainty **4**: 5-28.

HOFFMAN, ELIZABETH, KEVIN MCCABE, KEITH SHACHAT and VERNON L. SMITH (1994). "Preferences, Property Rights, and Anonymity in Bargaining Games." Games and Economic Behavior **7**: 346-380.

HOLT, CHARLES A. and SUSAN K. LAURY (2002). "Risk Aversion and Incentive Effects." American Economic Review **92**: 1644-1655.

JEDIDI, KAMEL and ROBERT E. KRIDER (1998). "Clustering at the Movies." Marketing Letters **9**: 393-405.

KAHNEMAN, DANIEL, JACK L. KNETSCH and RICHARD H. THALER (1986). "Fairness as a Constraint on Profit Seeking. Entitlements in the Market." American Economic Review **76**(4): 728-741.

KAHNEMAN, DANIEL and AMOS TVERSKY (1979). "Prospect Theory. An Analysis of Decision under Risk." Econometrica **47**: 263-291.

KONOW, JAMES (2000). "Fair Shares. Accountability and Cognitive Dissonance in Allocation Decisions." American Economic Review **90**(4): 1072-1091.

LADOUCEUR, ROBERT and MARIE MAYRAND (1987). "The Level of Involvement and the Timing of Betting in Roulette." Journal of Psychology **121**: 169-176.

MELLERS, BARBARA, ALAN SCHWARTZ, KATTY HO and ILANA RITOV (1997). "Decision affect theory." Psychological Science **8**(6): 423-429.

MELLERS, BARBARA, ALAN SCHWARTZ and ILANA RITOV (1999). "Emotion-based choice." Journal of Experimental Psychology: General **128**(3): 332-345.

MITZKEWITZ, MICHAEL and ROSEMARIE NAGEL (1993). "Experimental Results on Ultimatum Games with Incomplete Information." International Journal of Game Theory **22**: 171-198.

MNOOKIN, ROBERT H. and LEWIS A. KORNHAUSER (1979). "Bargaining in the Shadow of the Law." Yale Law Journal **88**: 950-997.

ROTHBART, MYRON and MARK SNYDER (1970). "Confidence in the Prediction and Postdiction of an Uncertain Outcome." Canadian Journal of Behavioral Science **2**: 38-43.

SIMONOFF, JEFFREY S. and ILANA R. SPARROW (2000). "Predicting Movie Grosses. Winners and Losers, Blockbusters and Sleepers." Chance **13**: 15-24.

SKINNER, SARAH J. (2006). "Estimating the Real Growth Effects of Blockbuster Art Exhibits. A Time Series Approach." Journal of Cultural Economics **30**: 109-125.

WANDTKE, ARTUR-AXEL and WINFRIED BULLINGER (2009). Praxiskommentar zum Urheberrrecht.

WATT, RICHARD (2006). "Licensing and Royalty Contracts for Copyright." Review of Economic Research on Copyright Issues **3**: 1-17.

Appendix A: Instructions Baseline

General Instructions for Participants

Welcome to our experiment!

If you read the following explanations carefully, you will be able to earn a substantial sum of money, depending on the decisions you make. It is therefore crucial that you read these explanations carefully.

During the experiment there shall be absolutely no communication between participants. Any violation of this rule means you will be excluded from the experiment and from any payments. If you have any questions, please raise your hand. We will then come over to you.

In any event, you will receive a lump sum of 2 euro for taking part in the experiment.

During the experiment we will not calculate in euro, but instead in Taler. Your total income is therefore initially calculated in Taler. The total number of Taler you accumulate in the course of the experiment will be transferred into euro at the end, at a rate of

1000 Taler = 1 Euro.

At the end you will receive from us the 2 euro plus the **cash** sum, in euro, based on the number of Taler you have earned.

Experiment Overview

The experiment consists of **8 rounds**, each of which has **4 stages** (maximum). In the experiment, there are 2 different roles, **Player S** (Seller) and **Player B** (Buyer).

At the beginning of the experiment, you are randomly allocated one of these two roles. During the entire 8 rounds of the experiment, you will remain in the same role.

At the beginning of each round, each Player S is paired with a Player B at random. In the course of the experiment, Player S **never plays twice** with the same Player B. And Player B, in the course of the experiment, **never plays twice** with the same Player S.

Stage 1: Player B makes Player S an offer.
Stage 2: Player S decides whether to accept or decline the offer.

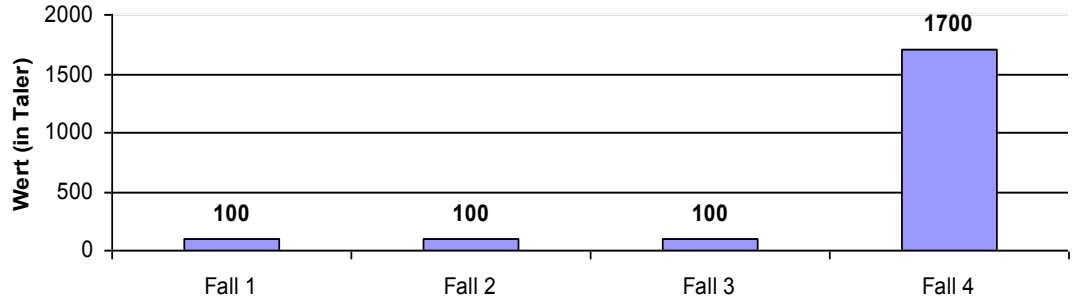
Stage 3: A draw decides whether the object is worth 100 or 1700 Taler.
Stage 4: Distribution of points.

Information on the Exact Proceedings of the Experiment

Stage 1:

Each player receives an **initial endowment** of 500 Taler onto a **Taler account**.

Player S is in possession of an object, Player B can purchase this object. **The object only has a value if Player B buys it**. At the time of purchase, however, the value of the object is still unknown. All you know is the distribution of these values.



In **3 out of 4** cases, a value of **100 Taler** is realized later.
In **1 out of 4** cases, a value of **1700 Taler** is realized later.

In order to purchase the object, Player B makes an **offer** to Player S.

Stage 2:

Player S hears about Player B's offer and decides whether to **accept** or **decline** Player B's offer.

If Player S **accepts** the offer, the object becomes Player B's possession, and the purchase price is transferred to Player S's Taler account (500 + purchase price). Player B's Taler account is reduced by the purchase price paid (500 – purchase price).

If Player S **declines** the offer, the round is ended. Stages 3 and 4 of this round are not played in that case. Both players' Taler accounts remain unchanged (500 Taler).

Stage 3:

A draw decides whether the object is worth 100 or 1700 Taler. You are told the result of this draw and of the account balances resulting from it.

- If the object is worth **100 Taler**, Player B's Taler account is: 500 – purchase price + 100.
- If the object is worth **1700 Taler**, Player B's Taler account is: 500 – purchase price + 1700.
- In both cases, Player S's Taler account is: 500 + purchase price

Stage 4:

Player B and Player S are given the opportunity to reduce the other player's Taler account by **distributing points**. Each point that you allocate to another player costs you **1 Taler** and reduces the other player's Taler account by **3 Taler**.

(At the latest) after Stage 4, the round is ended. Stage 1 of the next round follows, in which each Player S is paired with a new Player B and each Player B is paired with a new Player S. The Taler accounts of all players are saved (for the later payment) and reset to zero (for the new round).

After the 8 rounds, the Taler accounts saved after all 8 rounds are added up in order to calculate your payoff.

Appendix B: Control Questions Baseline

1. In Stage 1 Player B has offered Player S a purchase price of 10 Taler. In Stage 2 Player S has rejected the offer.

What is the
a) Income of Player B after Stage 2?.....
b) Income of Player S after Stage 2?.....

2. In Stage 1 Player B has offered Player S a purchase price of 150 Taler. In Stage 2 Player S has accepted the offer. In Stage 3 the random draw has determined the value of the commodity to be 100 Taler.

What is the
a) Income of Player B after Stage 3?.....
b) Income of Player S after Stage 3?.....

3. In Stage 1 Player B has offered Player S a purchase price of 80 Taler. In Stage 2 Player S has accepted the offer. In Stage 3 the random draw has determined the value of the commodity to be 1700 Taler.

What is the
a) Income of Player B after Stage 3?.....
b) Income of Player S after Stage 3?.....

4. After the random draw in Stage 3 (Price=90 Taler, Value=100 Taler) Player B has an income of 510 Taler and Player S of 590 Taler. In Stage 4, Player B allots 50 Points and Player S 0 Points.

What is the
a) Income of Player B after Stage 4?.....
b) Income of Player S after Stage 4?.....

5. After the random draw in Stage 3 (Price=150 Taler, Value=1700 Taler) Player B has an income of 2050 Taler and Player S of 650 Taler. In Stage 4, Player B allots 4 Points and Player S 200 Points.

What is the
a) Income of Player B after Stage 4?.....
b) Income of Player S after Stage 4?.....

Appendix C: Instructions Provision Treatment

Experiment Overview

We now repeat the experiment and introduce a few changes.

Once again, the experiment consists of **8 rounds**. Each of these rounds, however, no longer consists of 4 stages, but of (a maximum of) **7 stages**.

Stages 1, 2 and 3 are the same as in the first experiment. Stage 7 corresponds to Stage 4 of the first experiment. Stages 4, 5 and 6 are new.

There are now 3 different roles, **Player S** (seller), **Player B** (buyer) and **Player U** (umpire). At the beginning of the experiment, you are allocated one of the three roles at random. During the entire 8 rounds of the experiment, your role shall remain the same.

At the beginning of each round, a random procedure pairs each Player S with a Player B and a Player U. In the course of the experiment, Player S **never plays twice** with the same Player B. And Player B **never plays twice** with the same Player S in the course of the experiment.

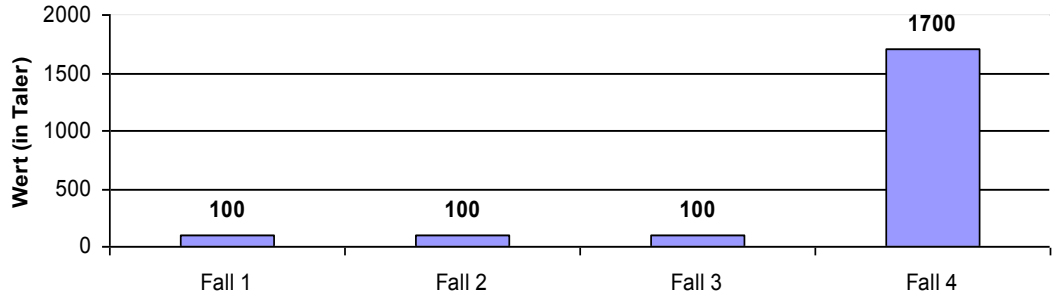
Stage 1: Player B makes an offer to Player S.
Stage 2: Player S decides whether to accept or decline the offer.
Stage 3: A draw decides whether the object is worth 100 or 1700 Taler.
Stage 4 (only if 1700): Player U names an appropriate purchasing price.
Stage 5 (only if 1700): Player B can make Player S a new offer.
Stage 6 (only if 1700): Player S decides whether to accept or decline the new offer.
Stage 7: Distribution of points.

Information on the Exact Proceedings of the Experiment

Stage 1:

Each player receives an **initial endowment** of 500 Taler, transferred to their **Taler account**.

Player S is in possession of an object, Player B can purchase this object. **The object only has a value if it is bought by Player B**. At the time of the purchase, however, the value of the object is still uncertain. You merely know the distribution of these values.



In **3 out of 4** cases, a value of **100 Taler** is realized later.
In **1 out of 4** cases, a value of **1700 Taler** is realized later.

In order to purchase the object, Player B makes Player S an **offer**.

Stage 2:

Player S is told the offer by Player B and decides whether to **accept** or **decline** Player B's offer.

If Player S **accepts** the offer, the object becomes Player B's possession, and the purchase price is transferred to Player S's Taler account (500 + purchase price). Player B's Taler account is reduced by the purchase price paid (500 – purchase price).

If Player S **declines** the offer, the round ends. Stages 3 to 7 of this round are not played in such a case. The players' Taler account remains unchanged (500 Taler).

Stage 3:

A draw decides whether the object is worth 100 or 1700 Taler. You are told the result of the draw as well as the account balances resulting from it.

- If the object is worth **100 Taler**, Player B's Taler account is: 500 – purchase price + 100.
- If the object is worth **1700 Taler**, Player B's Taler account is: 500 – purchase price + 1700.
- In both cases, Player S's Taler account is: 500 + purchase price.

Stage 4:

(Stage 4 is only played if the object is worth 1700 Taler.)

Player U decides which purchase price is to be deemed **appropriate**. This decision is initially not passed on to Player S and Player B.

Stage 5:

(Stage 5 is only played if the object is worth 1700 Taler.)

Player B now has the chance to make Player S a **new offer**. Player B may augment the puchase price that was paid to Player S in Stage 2, or reduce it, or leave it as it is.

Stage 6:

(Stage 6 is only played if the object is worth 1700 Taler.)

Player S is told Player B's new offer and decides whether to **accept** or **decline** Player B's new offer.

If Player S **accepts** the new offer, the purchase price paid in Stage 2 is no longer valid, and the new purchase price is transferred to Player S's Taler account.
Player S's Taler account is therefore: 500 + new purchase price.
Player B's Taler account is then: 500 – new purchase price + 1700.

If Player S **declines** the new offer, the purchase price paid in Stage 2 is no longer valid either. Player B and Player S are told which purchase price Player U deemed appropriate in Stage 4. This "**appropriate purchase price**" replaces the old purchase price.
Player S's Taler account is therefore: 500 + appropriate purchase price.
Player B's Taler account is then: 500 – appropriate purchase price + 1700.

Stage 7:

Player B and Player S are given the possibility of reducing the other player's Taler account by **distributing points**. Each point you distribute to the other player costs you **1 Taler** and reduces the other player's Taler account by **3 Taler**.

(At the latest) after Stage 7, the round ends. Stage 1 of the next round follows, in which each Player S is paired with a new Player B and a new Player U, and each Player B is paired with a new Player S and a new Player U. The Taler accounts of all players are saved (for the subsequent payoff) and reset to zero (for the new round).

After the 8 rounds, the saved Taler accounts from all 8 rounds are added up in order to calculate your payoff.

Appendix D: Control Questions Provision Treatment

1. After the random draw in Stage 3 (Price=150 Taler, Value=1700 Taler) Player B has an income of 2050 Taler and Player S of 650 Taler. In Stage 5, Player B confirms her offer from Stage 1 (150 Taler). In Stage 6, Player S accepts the new (=old) offer. In Stage 7, Player B allots 4 Points and Player S 200 Points.

What is the
a) Income of Player B after Stage 7?.....
b) Income of Player S after Stage 7?.....

2. After the random draw in Stage 3 (Price=150 Taler, Value=1700 Taler) Player B has an income of 2050 Taler and Player S of 650 Taler. In Stage 5, Player B increases her offer to 400 Taler. In Stage 6, Player S accepts the new offer. In Stage 7, Player B allots 0 Points and Player S 0 Points.

What is the
a) Income of Player B after Stage 7?.....
b) Income of Player S after Stage 7?.....

3. After the random draw in Stage 3 (Price=150 Taler, Value=1700 Taler) Player B has an income of 2050 Taler and Player S of 650 Taler. In Stage 5, Player B increases her offer to 747 Taler. In Stage 6, Player S rejects the new offer. The Players B and S are informed that Player U regards the appropriate purchase price to be 50 Taler. In Stage 7, Player B allots 0 Points and Player S 0 Points.

What is the
a) Income of Player B after Stage 7?.....
b) Income of Player S after Stage 7?.....

4. After the random draw in Stage 3 (Price=150 Taler, Value=1700 Taler) Player B has an income of 2050 Taler and Player S of 650 Taler. In Stage 5, Player B increases her offer to 583 Taler. In Stage 6, Player S rejects the new offer. The Players B and S are informed that Player U regards the appropriate purchase price to be 950 Taler. In Stage 7, Player B allots 250 Points and Player S 100 Points.

What is the
a) Income of Player B after Stage 7?.....
b) Income of Player S after Stage 7?.....

Chapter 2

The Coevolution of Behavior and Normative Expectations:
Customary Law in the Lab

Christoph Engel and Michael Kurschilgen

Abstract

Customary law has been criticized from very different angles. Rational choice theorists claim that what looks like custom is nothing but self-interest. Positivists doubt that anything beyond consent assumes the force of law. In this paper, we adopt an experimental approach to test these claims. We show that both critics miss an essential feature of custom. Custom guides behavior as normative expectations and behavioral patterns coevolve. Customary law capitalizes on this more general social mechanism. If it is not backed up by sanctions, customary law is not more effective than mere custom in realigning individual action and social well-being. Yet if the emerging rule is perceived to be grounded in law, sanctions and custom are complements, while they are substitutes otherwise.

Keywords: Customary Law, Normativity, Crowding Out, Public Good, Experiment

JEL: C14, C91, D03, D62, D63, H41, K10

I. Introduction

Conceptual rigour can be a dangerous weapon. From several directions, this weapon has been directed to customary law, with lethal effect as aggressors pretend, or leaving the target badly wounded but able to recover as defenders believe. This article objects that neither position gives customary law the credit it deserves. It uses a laboratory experiment to show the power of customary law, and the behavioural forces driving it, thereby contributing to the nascent experimental law and economics literature (characteristic contributions include Croson and Johnston 2000**;** Arlen, Spitzer et al. 2002**;** Loewenstein and Moore 2004**;** McAdams and Nadler 2005**;** Nadler and Seidman Diamond 2008**;** Grechenig, Nicklisch et al. 2010**;** Zeiler 2010).

In almost all legal orders of the world, customary law is acknowledged as a valid source of law. Law need not originate in legislation or precedent, but may be created by the behaviour and the will of those supposed to abide by it. Yet in most national legal orders, customary law in the strict sense has become rare. Within nation states, if there is need for a new rule, more convenient law making procedures are readily available. Society need not wait until custom has formed. Moreover, when a rule is generated by explicit decision in a formalized procedure, this notably reduces ambiguity about its precise contents. In contrast, substitutes for customary law are less easily employed in the dealings of sovereign states with each other, the main reason for this difference being, the absence of a sovereign ruler who could ordain reluctant states to subdue to the common will or good. This explains why, these days, most of the legal debate on customary law is conducted by international lawyers.

In this paper, we use a lab experiment to test the main claims made in the legal debate over customary law. This creates an obvious tension. The debate originates in a field of law where entire states are the main actors. We test these claims with individuals. Since it is impossible to bring states to the lab, or to engage them in a field experiment, we have to trade off a loss in context specificity for a gain in experimental control and thus causal inference. We do of course not mean to argue that states essentially behave the same way as individuals. States are highly aggregate, institutionally and historically embedded corporate actors. We abstract from all of this. Yet we believe this price for experimental control to be justified for three reasons: First, the legal debate does not posit that customary international law is conceptually different from customary law in national law. Insights that are valid for customary national law, which does engage and address individuals, are therefore in

principle also valid for international law. Second, the arguments brought forward in the international law debate over customary law, which we review in the next section, do not rely on the character of states as corporate actors, but invoke mechanisms that hold for any actor, and for individuals as well. Third, a very similar legal debate is indeed concerned with the normative force of custom on the behaviour of individuals. While this is not customary law in the strict sense, there are many instances where the law acknowledges the normative relevance of custom, short of regarding it as a source of law. Classic illustrations are *lex mercatoria,* trade practice, or codes of conduct.

We argue that the critics of customary law use an overly narrow and therefore inappropriate concept of normativity. Norms do not only direct behaviour if norm violation is against the addressee's self-interest. On the other hand, the duty to abide by the law in force is not the exclusive motivating force either. Norms matter because they provide guidance. Most actors are most of the time willing to follow the norms prevailing in their context, or at least to be not too far off the mark, and most actors expect other actors to be thus guided.

The article takes issue with the furthest reaching claim: customary law is "epiphenomenal", i.e. what looks like an effect of law actually is nothing but an act of self-interested behaviour. The proponents of this claim rely on game theory. They model states as unitary actors, i.e. as if they were individuals. In the experiment, we test a situation where game theory unequivocally predicts total defection: a public good game. In line with a rich literature in experimental economics, we refute the claim even in our *Baseline*, where normativity plays no (explicit) role. In our *Law* treatment, we introduce a meta-rule for the formation of customary law. In game-theoretic terms, this meta-rule is totally irrelevant. Yet it turns out to have a pronounced positive effect on people's cooperativeness.

Positivists will not be surprised by this result. Since we have explicitly invoked the legal order, this is what they would expect. Yet for them, the motivational force rests in the legal order. Actors abide by the law since this is their duty. Consequently, if we do not invoke the law but only ask participants whether they believe a (non-legal) norm to exist, the effect should vanish. This is what we test with our *Comity* treatment. Contrary to the positivist prediction yet in line with the position of Maurice Mendelson (Mendelson 1998), this turns out to be at least as effective at enhancing cooperation as the *Law* treatment.

Based on these findings, we propose an alternative conceptualisation of customary law: Customary law guides behaviour into the normatively desired direction as normative expectations and behavioural patterns coevolve. We back this claim by further data analysis. What our participants have (privately) stated in the preceding period significantly explains their behaviour in the subsequent period; what the group has done in the preceding period significantly explains their statements about norm existence and minimum required contributions in the subsequent period.

Thus far, our results seem to suggest that the power of customary law boils down to the power of normativity. Law would be immaterial, not because it has no behavioural effect, but because one does not need the legal order to bring the effect about. To test this hypothesis, we rerun the experiment in a setting where participants can sanction each other *(Baseline-S)*. Also in this setting, the introduction of the meta-rule for the formation of customary law (treatment *Law-S*) has a strong positive effect on cooperation compared to *Baseline-S*. However, if we only privately ask participants whether there is a norm (treatment *Comity-S*), cooperation is even lower than in *Baseline-S*. If participants do not perceive the norm to be legal, sanctions "crowd out" some of the beneficial effect of the norm. By contrast sanctions and normative expectations corroborate each other if normative behaviour is required in law.

In the next section, we develop the legal research question from the lively debate over customary law in public international law, and derive the hypotheses to be tested in the experiment. Section 3 presents the design of our first experiment. Section 4 reports the results. Section 5 investigates the driving forces. Section 6 reports findings from the additional three treatments with sanctioning opportunities. Section 7 concludes.

II. The Legal Debate

In an influential, provocative paper, Jack Goldsmith and Eric Posner have argued that customary international law is a mere epiphenomenon. What looks as if states were abiding by international law effectively is nothing but an exercise of self-interest, they say. States maximise their utility. They cooperate if this is the optimal strategy, given the (expected) behaviour of other states. Cooperation may even occur if states face a social dilemma. Through the folk theorem, if their interaction is repeated and the end is uncertain, cooperation may occur as long as neither of them is too

impatient (Aumann and Shapley 1994). Yet cooperation is much less likely to occur if the group is large. In an n-person prisoner's dilemma, the conditions for cooperation are close to heroic (Goldsmith and Posner 1999).

> "A nation's 'compliance' with the cooperative strategy in the bilateral prisoner's dilemma has nothing to do with following a norm from a sense of legal obligation. Nations do not act in accordance with a norm that they feel obliged to follow; they act because it is in their interest to do so" (Goldsmith and Posner 1999**:** 1132).

Many have taken issue with this claim (for a survey see Norman and Trachtman 2005). Critics have in particular argued that the authors give too little credit to reputation, retaliation and segmentation (Chinen 2001), that international law changes the payoffs of the game (Guzman 2008), and that the strategies of grim trigger and penance make it possible to sustain cooperation even if the number of actors is large (Norman and Trachtman 2005). Our approach differs in that we take exactly a situation Goldsmith and Posner claim makes cooperation impossible, and test experimentally whether it nonetheless occurs. We thus test:

H_1: In a multi-person dilemma game of finite horizon, there is no cooperation. A meta-rule requiring persons to abide by the rules of customary law is immaterial.

The positivist tradition stands in sharp contrast to the ideas of Goldsmith and Posner. For them the force of customary law rests in the individual addressee's consent, either to a specific rule on the issue at hand (Triepel 1899**;** Anzilotti 1955), or to a meta-rule stipulating the conditions under which a new rule comes into being (Kelsen 1952**;** Morelli 1967); (Elias 1995). Consequently, the scope for customary law that positivists are willing to grant is rather small.

Positivists are not primarily interested in extra-legal effects. Their main contribution is to the doctrine about rules on rules. Positivists define the conditions under which an utterance of words assumes the force of law. Yet the exercise has a natural corollary. If the law's addressees care about the law at all, they should care much more if this utterance of words is actually law, rather than a mere statement about desirability, on whatever non-legal grounds.

From this angle, the positivist position is related to a facet of the debate over an "expressive" function of law. While some contributions narrowly conceive law as a mere sanction (Bohnet and Cooter 2001**;** Tyran and Feld 2006**;** Galbiati and Vertova 2008**;** Bernasconi, Corazzini et al. 2010), others adopt a richer concept of law, arguing that the law serves as a focal point (McAdams 2000**;** McAdams and Nadler 2005**;** McAdams and Nadler 2008), informs people about behaviour others will approve (McAdams 2000), induces people to change their beliefs about the consequences associated with an action (Geisinger 2002), or changes the perception of underlying social norms (Feldman and Nadler 2006). Empirical findings have been mixed. The willingness to disregard the copyright protection of electronic works was not significantly affected by either making the illegality salient, nor by also pointing to informal sanctions, like a loss in reputation (Feldman and Nadler 2006). By contrast, in Switzerland voter turnout went down once a canton had abolished the legal obligation to vote, although enforcement had only been symbolic (Funk 2007).

As long as there is no enforcement, in psychological parlance the difference between mere comity and actual law boils down to a frame. It makes people see the issue in a different light. The requested behaviour is not only desirable; it is mandated by the legal order. Frames have been shown to strongly influence behaviour. Whether context makes people see a choice as involving a gain or a loss has a heavy impact, despite the fact that the task can easily be reframed by manipulating the reference point (Tversky and Kahneman 1981). Also if a choice is contrasted with another, irrelevant outside option, choices change substantially (Tversky and Simonson 1993). When an agency aims at eliciting voters' willingness to pay for a public venture, responses heavily depend on how the issue is presented (Kahneman, Ritov et al. 1999). In all these tasks, the frame activates people's world knowledge. By the same token, we expect that people's attitudes toward law in general are activated if it is made salient that custom can be binding law (for a more elaborate theory of what this implies see Engel 2008). Most legal rules are meant to tame egoism and to make the law's subjects see the issue in the light of what is socially desirable. We therefore derive the following positivist prediction:

H_2: People are more likely to overcome a social dilemma if they are made aware that this is their duty in law.

Rational choice theorists have a hard time with customary law since it seems unclear why self-interested actors should contribute to the formation of a rule that will

prevent them from acts of selfishness. Positivists have a hard time with customary law since it seems unclear how a new norm of customary law could ever come into being. Must those who originally claim the norm to exist actually have been mistaken (Elias 1995**:** 503; Stern 2001**:** 97)? Are the proponents of a new rule of customary law actually only making a proposal, which must be accepted by other actors (Elias 1995**:** 508)? Must one have recourse to some external authority, like divine emanation, natural law, or social necessity, to explain the formation of a new rule of customary law (Stern 2001**:** 92)?

It seems that both the positivist and the rational choice conceptualisations of customary law miss a key ingredient. Practice turns into law since behaviour and normative expectations coevolve. If nearly everybody behaves in a certain way, this shapes beliefs about others' future behaviour. If the pattern has been repeated for a while, the *behavioural* belief (i.e. how others will probably behave) turns into a *normative* expectation (i.e. how others should rightly behave). The belief is no longer is purely cognitive. A motivational component is added to it. If an actor deviates from established practice, she violates others' normative expectations. Others regard such behaviour not only as anti-social but as illegitimate.

In his Hague lectures, Maurice Mendelson has offered a related explanation (Mendelson 1998). For him, the textbook approach to customary law misses a central feature of public international law. The international legal order is "semi-anarchic" (166), embryonic, and in a deep way incomplete. "Whilst modern domestic societies are characterised by highly centralised and compulsory systems of law-making and adjudication, not to mention enforcement, international society is not like that" (168). Therefore a "formalistic approach" (168) is misplaced. It is not possible to state in an abstract way the conditions that must be fulfilled for a new rule of customary international law to come into being (172). "The characteristic of this kind of law is that it is not just unwritten, it is informal" (172). The customary process is in fact a continuous one, which does not stop when the rule has emerged […]. Even after the rule has 'emerged', every act of compliance will strengthen it, and every violation, if acquiesced in, will help to undermine it" (175). Customary international law rests on the conviction that "states should comply with the legitimate expectations of the international community" (185), where the ambiguity of the term "expectation" is deliberate: "If, within a social group, people habitually behave in a certain way, then, particularly if others rely on the continuation of this conduct, the sentiment may develop within that society that one is obliged to continue so to act. In other words, a

norm emerges from what is normal […]. If the generality of states has regularly behaved in certain ways […], then a legitimate expectation arises that they will continue to do so" (185 f.).

In essence, this is a claim about normativity in general. Yet it can be combined with the reasons expressive law theorists give for the specific behavioural effects of law, which we have reported above. If this were to hold true, we would have to see the following:

H_3: Normative expectations and behaviour coevolve. Coevolution is faster and more robust if the normative expectation originates in the legal order.

III. Experimental Design

To test these hypotheses, we run a standard linear four person public goods game. In this game, payoffs are given by (1):

$$\pi_i = e - c_i + m\sum_{j=1}^{N} c_j \qquad (1)$$

Every period, each participant receives the same endowment e. She can freely decide how much of the endowment she wants to keep and how much she wants to contribute c_i to a joint project. Contributions by all members are multiplied by the marginal per capita rate $m < 1$. As long as $mN > 1$, the society of all group members is best off if everybody contributes everything. However individually, each member is best off if others contribute while she freerides. This constitutes an n-person prisoner's dilemma with continuous action space. As is standard in the experimental literature on public goods (for overviews see Ledyard 1995**;** Zelmer 2003**;** Chaudhuri 2011), subjects are informed after how many periods the game ends. Under standard game theoretic assumptions, players reason backwards. Since it is rational to defect in the last period, to preempt being the sucker in this period, a rational player defects in the penultimate period, and so forth until the first period (Selten 1978**;** Rosenthal 1981). Hence the unique subgame-perfect Nash equilibrium is all players defecting from the beginning. Note that, by this design, we give the rational choice critique of customary law its best shot. Since the number of periods is announced and subjects

interact anonymously, game theory would even predict that customary law is pointless if there were only two players.

In our *Baseline* treatment participants play this game in fixed groups of 4, interacting over 30 announced periods, with an endowment of 20 experimental currency units (ECU), and a marginal per capita rate $m = 4$. These parameters are standard in the experimental literature on public goods, except for the number of periods. We have replaced the usual duration of 10 periods by a longer spell since we want more scope for analysing the coevolution of norms and behaviour. With these parameters, if all participants contribute their entire endowments to the joint project, all receive 32 ECU. If all keep their entire endowments, all receive 20 ECU. If three participants contribute fully, while the fourth keeps everything, the former have 24 ECU, while the latter has 44 ECU. If one contributes fully, while three freeride, the former has 8 ECU, while the latter each have 28 ECU.

The first stage of the *Comity* treatment is exactly the same as the baseline. However, *Comity* has one additional stage: In each period, after participants have made their contribution decisions, but before giving them feedback, we ask them the following two questions:

> "1. Do you believe in your group exists a general norm regarding an adequate minimum contribution to the project (yes/no)?
> 2. If so, which is the generally expected minimum contribution (number from 0 to 20)?"

The instructions make it clear that participants will not get feedback about other participants' statements.

The *Law* treatment is identical to *Comity* but for the fact that, in addition, participants read the following paragraph in the instructions:

> "For new law to originate, it is not necessary that the legislator pass a statute, or that the parties agree on an explicit contract. Customary law is equally valid and binding. Customary law comes into being if the large majority of those affected for a sufficiently long period behave in a sufficiently similar way. The fact that some contribute even more to the joint project does not prevent that a norm of customary law originates. Hence a rule of customary law may

prescribe a minimum standard. Customary law may originate here in the lab as well."

Note that we have deliberately kept this paragraph procedural. Participants are informed about the conditions under which *they themselves* are able to make new law. This is not only in line with customary law doctrine, but also a safeguard against the risk of imposing any specific material norm on participants.

The experiment was run in the Bonn EconLab and programmed in zTree (Fischbacher 2007). Participants were invited using ORSEE (Greiner 2004). After being seated in individual cubicles, participants received experimental instructions and answered a set of control questions (see the Appendix). We had 20 participants, interacting in 5 groups in the *Baseline*,[1] and 24 participants, interacting in 6 groups, in each of the two treatments. Participants were randomly drawn from a pool of some 3500 subjects. They held various majors. Approximately half of them were female. In the *Baseline*, participants on average earned 12.42 € (15.73 $), in the *Comity* treatment earnings were on average 14.07 € (17.82 $), and in the *Law* treatment 13.80 € (17.48 $).

IV. Results

A. Treatment Effects

From Figure 1 one directly sees that both our treatments have a pronounced effect on cooperation. Whereas the *Baseline* displays the characteristic decay of contributions over time (see for example Fehr and Gächter 2000), both in *Comity* and in *Law*, average contributions even increase in the first periods, and they are much higher than in the *Baseline* until the endgame effect kicks in. Descriptively, *Comity* has a slightly stronger effect on contributions than *Law*.

[1] The planned 6th group could not be filled since participants did not show up.

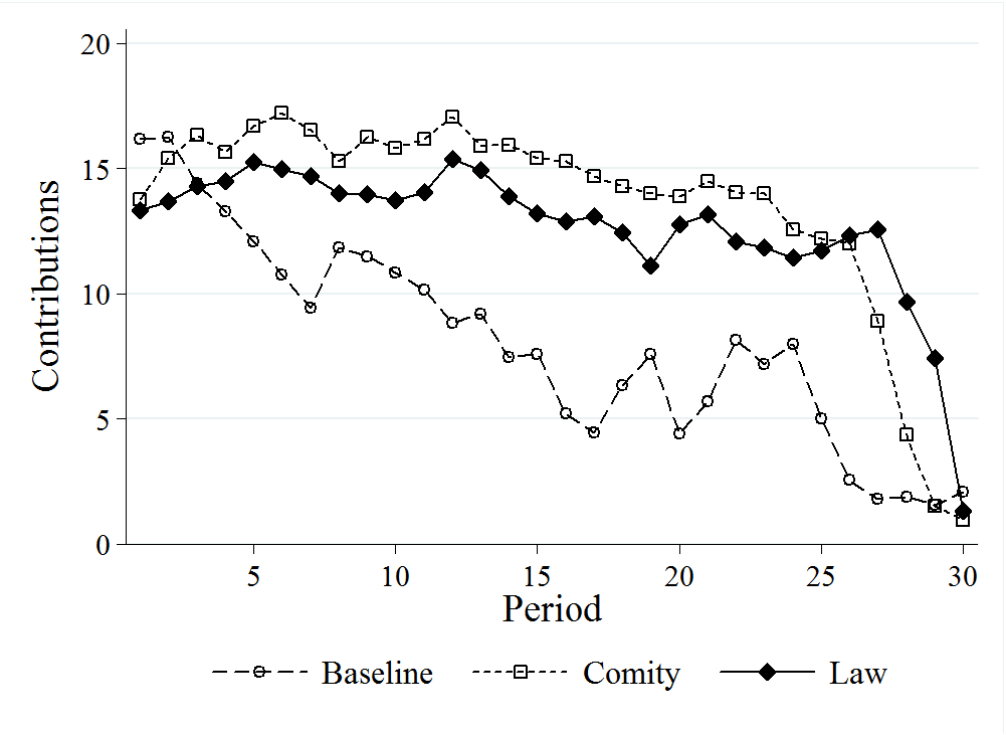


Figure 1: Contributions to the Public Good
Note: Cumulative distribution function, displaying the share of individuals willing to pay (up to) a certain price to deviate from material selfishness. Every deviation from the selfish optimum implied forgone payoffs of 10 tokens.

The visual impression is confirmed by statistical analysis. In a conservative non-parametric test over means per group, the difference between the *Baseline* and *Comity* is significant at conventional levels (Mann Whitney, N = 11, p = .0446). In this test, the difference between the *Baseline* and *Law* is insignificant (p = .1441). We analyse the data parametrically using a random effects Tobit estimator.[2] Using this procedure,

[2] The random effect captures the dependence at the level of individuals and the Tobit functional form accounts for the fact that our dependent variable is both left and right censored (many participants contribute their entire endowment of 20 tokens, while many contribute nothing). Participants interact in the same group of four over 30 periods. This creates a second source of dependence. Ideally we would therefore want to estimate a mixed effects model with two random effects: one for the group and another for the individual. Yet unfortunately there is no generally acknowledged mixed effects estimator for censored data. As a substitute, to make sure standard errors are not deflated, we bootstrap the estimator, with random draws of entire groups.

in the regressions of Table 1 we establish a significant treatment effect for both *Comity* (model 1) and *Law* (model 2), in comparison with the *Baseline*. By contrast the difference between *Comity* and *Law* is clearly insignificant (model 3, p-value *Law* = .535).

Our results clearly refute $\mathbf{H_1}$, which was derived from the rational choice critique of customary law. In a way, already our *Baseline* speaks against the claim that there is no cooperation in a multi-person dilemma. For sure, in the reality of international law, there are many more than just four actors. Yet note that a four person game excludes stabilising cooperation by threatening a defector with Nash reversion. Since there is more than one partner, players cannot use their own contributions, in the subsequent period, as a sanctioning technology. They would not only hit free riders, but also those who have faithfully contributed to the joint project. On the conceptual grounds on which the rational choice critique of customary law is built, a four actor dilemma is no different from a 192 actor dilemma.

Moreover, as model 2 shows, if we only compare the *Baseline* with the *Law* treatment, law clearly matters. The fact that behaviour is required by law has a big, positive effect. Note that our *Law* treatment is very subtle. The legal norm not only lacks sanctions, there is not even communication among those expected to abide by the legal rule, neither about its existence nor about its contents.

Table 1: Treatment Effect on Contributions

Dependent Variable: Contributions to PG

	(1) Baseline vs. Comity	(2) Baseline vs. Law	(3) Comity vs. Law
Comity	13.661*		
Law		9.597*	-3.150
Period	-.813***	-.604***	-.517***
Cons	18.175***	15.217***	26.483***
N	1320	1320	1440
left censored	311	327	198
right censored	469	380	595
p model	<.001	<.001	<.001

Note: random effects Tobit, ll(0) ul(20), bootstrapped with random draws of entire groups. 50 reps. reference category: models 1 and 2: *Baseline*, model 3: *Comity*. *** p < .001, ** p < .01, * p < .05

By contrast, we do not find support for **H_2**, which we derived from legal positivism, and from expressive law theory. We do not find any statistical difference between *Comity* and *Law*. In our data, customary law boils down to normativity. The fact that the rule is embedded in the legal order does not help participants overcome the dilemma even better. Descriptively they even fare slightly worse. We will revisit this issue in our second experiment, but first turn to our process hypothesis **H_3**.

B. The Coevolution of Behavior and Normative Expectations

Figure 2 shows for both the *Comity* and the *Law* treatment that normative expectations and behaviour are indeed synchronous. As long as many participants believe there is a norm, contributions are high. If this belief erodes, ultimately behaviour is affected. There is also a clear relationship between statements about the minimum expected contribution and actual contributions.[3]

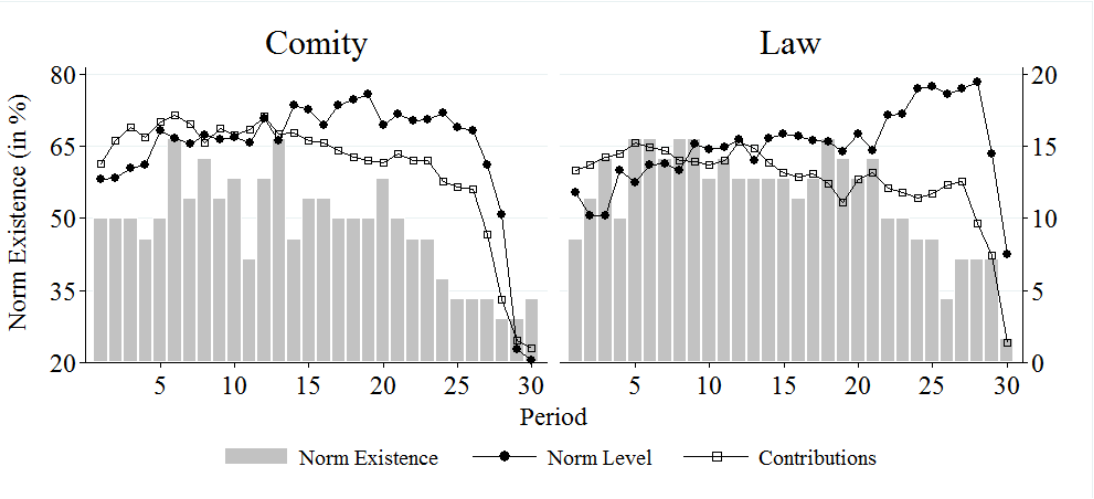


Figure 2: Coevolution of Behaviour and Normative Expectations
Note: Left axis denotes % of participants who say there is a norm. Right axis denotes actual and expected contribution.

[3] The apparent kink in statements about the required contribution level by the end in the *Law* treatment is due to a selection effect. If we analyse first differences of these statements, in neither treatment do they go up over time. Hence in the final periods, those remaining faithful to the norm do not believe it to be even stronger. Rather those who always believed the norm to be more demanding are more likely to still think it exists.

We can exploit the panel structure of our data to identify effects in both directions. If a participant has claimed in the previous period that there is a norm, this significantly and substantially increases how much she contributes to the joint project in the subsequent period (Table 2, model 1). Likewise, if one only looks at those participants who claim norm existence, the higher a participant has claimed the minimum contribution to be, the more she contributes in the subsequent period (model 2). This finding suggests a desire for self-consistency between one's stated normative expectations and one's contribution behaviour. In the reverse direction, the higher mean contributions in the participant's group in the previous period[4], the more this participant is likely to state that there is a norm (model 3).[5] Likewise, the higher mean contributions in the previous period, the more demanding are the statements about the required minimum contribution (model 4).

Table 2: Coevolution of Behaviour and Normative Expectations

	(1)	(2)	(3)	(4)
Dependent Variable:	contribution	contribution	statement of norm existence	statement of norm level
Lagged statement of norm existence	6.465**			
Lagged statement of norm level		1.699***		
Lagged mean contribution in group			.202***	1.189***
Law	-3.530	-1.231	.784	-.914
Period	-.510***	-.445*	-.030**	.123^{+}
cons	23.408***	3.803	-2.588***	-3.583
N	1392	724	1392	715
left censored	198	54		23
right censored	574	424		303
p model	<.001	<.001	<.001	<.001

Note: models 1,2,4: random effects Tobit, ll(0) ul(20). Standard error from bootstrap, sampled at group level, 50 reps. model 3: mixed effects logit, period nested in individual nested in group. *** p < .001, ** p < .01, * p < .05, $^{+}$ p < .1.

[4] We work with the lag of average contributions, although participants made their statements after contribution choices since they did not yet have feedback about this period's contributions.

[5] In a logit model, both events are equally likely if the predicted coefficient is 0.

Figure 3 separately draws the contributions of participants who affirm and who deny the existence of a norm, together with the lagged mean contribution of all participants. The left panel displays the *Comity* treatment, and the right panel the *Law* treatment. As one sees, in all periods those who claim norm existence on average contribute more than the mean in the previous period. By contrast, those who say there is no norm contribute less than the lagged mean in all but the very first periods. Moreover, until the endgame effect kicks in, contributions of those who say there is a norm are almost stable, while contributions of the remaining subjects visibly decay over time. We thus support the main claim of **H_3**: normative expectations and behaviour coevolve. These observations hold likewise for the *Comity* and the *Law* treatment. We do however not find any additional effect of law over mere social expectations. The next section further examines this result within a slightly modified experimental paradigm.

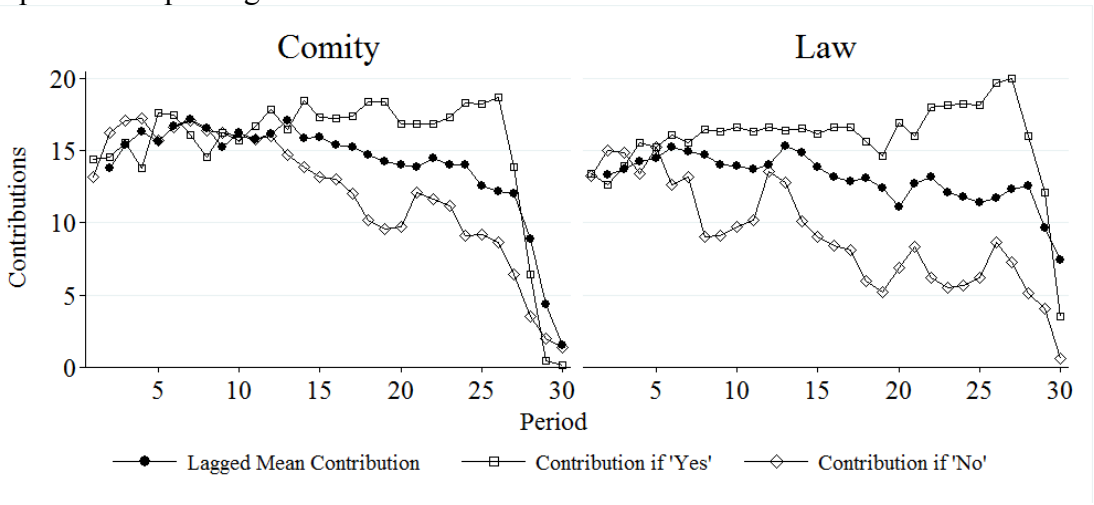


Figure 3: Contributions by Claim of Norm Existence

C. Law beyond Normativity?

The results reported above strongly suggest that customary law matters. It substantially improves behaviour, as we have clearly seen in Figure 1. This is welcome news for public international law, and for law in general. Yet it is troubling that, seemingly, an equally strong behavioural effect could be brought about without explicit reference to law. Seemingly, customary law only matters because it makes normativity salient. In the terminology of international law doctrine we might thus

conclude that customary law is not more effective than mere comity. If that was the end of the story, why has (customary) law evolved? Why do states seem to care? Why do they exert considerable effort to show that what might appear to be in violation of customary international law actually is in harmony with it, for instance because the rule has an exception? Within international law doctrine, the answer is straightforward: If and only if the state that has violated an obligation from customary international law, it must tolerate that the victim retaliates by itself violating another rule of international law (within the limits of the law of reprisals, for sure). Yet is this doctrinal distinction of any behavioural relevance?

To answer these questions, we have run a second experiment. In this experiment we repeat our three treatments from above with a slight modification. In every period we now add an additional stage in which participants have the possibility to sanction each other (cf. Fehr and Gächter 2000**;** Nikiforakis and Normann 2008). Specifically, to destroy one ECU of another group member, a participant must spend one ECU of her own period income. The first stage payoff function remains as in (1). If we write π_1 for the first stage income, total payoff is given by (2):

$$\pi = \pi_1 - \sum_{j \neq i}^{N} s_{ij} - \sum_{j \neq i}^{N} s_{ji} \qquad (2)$$

First stage income is reduced by all sanctioning points this player gives to another group member (s_{ij}), and by all sanctioning points other group members inflict on her (s_{ji}). Another 72 students have participated in these three treatments with sanctioning opportunities ("S-treatments"). They have on average earned 14.45 € (18.44 $) in the *Baseline-S*, 13.78 € (17.58 $) in *Comity-S*, and 15.19 € (19.38 $) in *Law-S*.

Of course, the caveat made when introducing the first experiment applies here too: we test students, and the reactions of states to the availability of a sanctioning mechanism may well be driven by other forces. In particular, the cost or risk involved in enforcing international law against a powerful state may be very high. Yet note that apart from that our design is very close to the situation of customary international law: There is no central enforcement agency. The cost of enforcement is borne by those who engage in it. If the rule is valid *erga omnes* (for an economic analysis see Engel 2009**;** Posner 2009), each state is entitled to enforce it, even if another state has been violated. Since the rule originates in state practice, it need not be explicit. Frequently, whether there is a rule or not is disputed. Also, using a 1:1 fine-to-fee

ratio, we have made punishment as costly as it reasonably can be made in the lab; typically in public good experiments the leverage of punishment is much higher.

Figure 4 compares cooperation in the three treatments without (left panel) and with (right panel) decentral sanctioning. There are two messages. The first is straightforward. Sanctions are very effective at improving cooperation, even if they are weak. For the rational choice theorists of customary international law this again should be surprising. In our game every player is best off if others bear the cost of sanctioning, which is why theory predicts zero sanctions; the original dilemma repeats among those actors who are willing to contribute to the public good themselves, but would rather have other loyal contributors bear the cost of enforcing it (Yamagishi 1986**;** Heckathorn 1989). Consequently adding the sanctioning option should not change behaviour. Yet empirically people, and states for that matter, are willing to engage in costly sanctioning. Potential addressees rightly expect sanctions, and react by changing their behaviour (key contributions to this literature are Fehr and Gächter 2000**;** Fehr and Gächter 2002). Both the difference between *Comity* and *Law-S* ($N = 12$, $p = .0163$) and the difference between *Law* and *Law-S* ($N = 12$, $p = .0104$) is highly significant in a simple Mann-Whitney test over group means.

The more important message of Figure 4 is more subtle, but also more relevant for understanding the difference between comity and customary law. While contributions were slightly lower in *Law* than in *Comity* when there was no sanctioning option, in the presence of sanctioning the order reverses. Now *Law-S* outperforms all other treatments, while *Comity-S* performs even poorer than the *Baseline-S*. There is a well-understood behavioural explanation for the latter effect. Extrinsic interventions crowd out intrinsic motivation (Bolton and Katok 1998**;** Fehr and Gächter 2001**;** Frey and Jegen 2001**;** Nyborg and Rege 2001**;** Chan, Godby et al. 2002**;** Janssen and Mendys 2004**;** Eckel, Grossman et al. 2005**;** Borges and Irlenbusch 2007). Yet interestingly the socially detrimental effect of sanctions disappears if behaviour is not only socially but legally required. Now, to the contrary, sanctions and intrinsic motivation are no longer substitutes. They become complements. Law makes sanctions behaviourally robust.

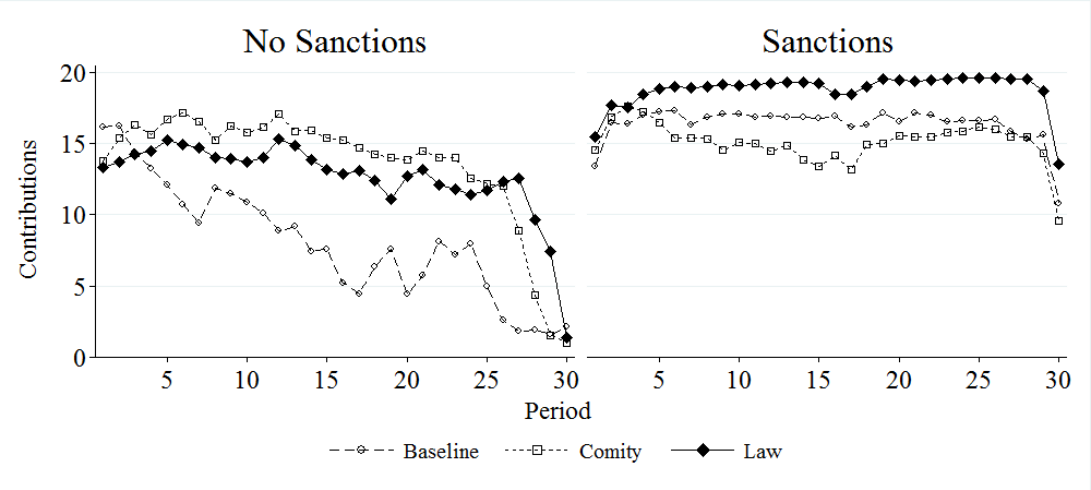


Figure 4: Contributions with and without Sanctions

Table 3 analyses the complementarity between law and sanctions in further depth. Specifically, we want to see how people change their contribution behaviour as a result of receiving a sanction in the previous period. Consequently, we do not look at contribution levels, but at contribution changes, i.e. first differences.

Table 3: Sensitivity to Sanctions

lagged received sanction	.659***
Comity	-.161
lagged received sanction*Comity	-.217**
period	-.008
period 30	-5.418***
cons	-.026
N	1392
p model	<.001

Note: Dependent variable: first differences of contributions. Mixed effects model, period nested in individual nested in group. Reference group: *Law* treatment. *** p < .001, ** p < .01, * p < .05, $^{+}$ p < .1.

There is a strong main effect of sanctions: each sanctioning point a participant has received in the previous period induces her to increase her contributions by .659 points. More importantly even, from the negative interaction of sanction with *Comity-S* we see that the beneficial effect of punishment is significantly less pronounced when the normative expectation does not originate in law. In the *Comity-S* treatment,

one punishment point only induces the recipient to adjust her contributions by .659 - .217 = .442 points.

Figure 5 provides additional support for our process hypothesis **H3**. In addition, one clearly sees that the main difference between *Comity-S* and *Law-S* originates in the perception of norm existence. If we make the possibility of the formation of customary law salient, and if participants have the possibility to sanction norm violations, after a small number of periods almost all participants believe a norm to exist. And this norm closely corresponds with actual behaviour. One way of explaining this effect is game theoretic. Our *Law-S* treatment can be interpreted as turning the meta-rule, i.e. the possibility that a normative expectation emerges, into perceived common knowledge.[6] Our experiment thus also points to an additional option for explaining the effect of customary law when keeping the rational choice assumption that actors just maximise their payoffs.

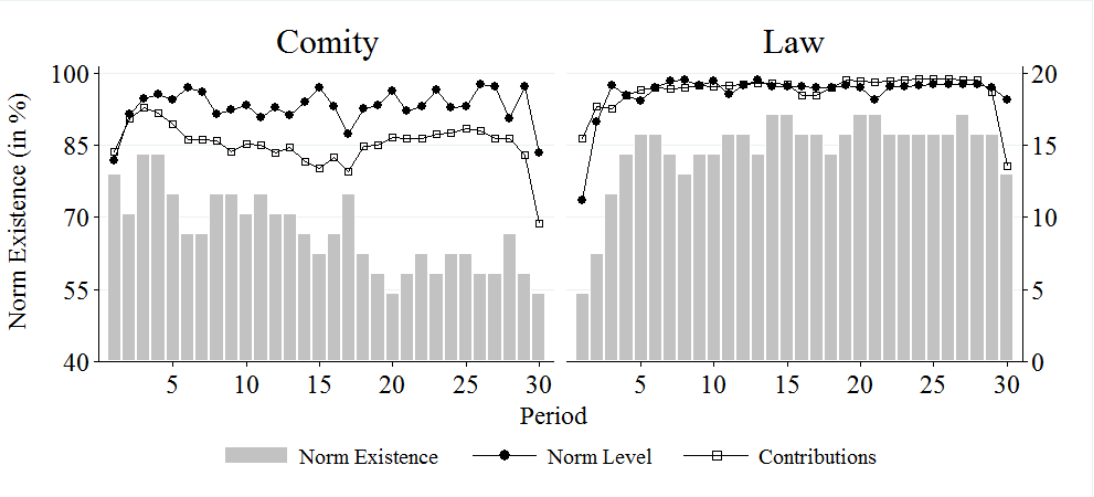


Figure 5: Norm Existence and Norm Level with Sanctioning

Figure 5 shows that the dynamics of norm perception are substantially different in *Comity* and *Law*. The positive and significant time trend indicates that participants become more and more likely over time to state there is a norm when this norm originates in law. However, they are less likely to say so in the absence of law (.100 -

[6] We are grateful to Carlos Alós-Ferrer for this interpretation. Descriptively, we already see a similar effect without sanctions, Figure 2, yet it is not significant.

.217 = -.117). This qualifies our finding from the first experiment. While we do not find support for **H_2** in the absence of sanctions, once the law is backed up by a sanction, law is indeed more powerful than mere comity. In our second experiment we thus support **H_2**.

Table 4: Statement of Norm Existence with Sanctions

Comity	2.059
Period	.100***
Comity*period	-.217***
Cons	2.280*
N	1440
p model	<.001

Note Dependent variable statement of norm existence (yes/no). Mixed effect logit, statement nested in individual nested in group. Data from *Comity* and *Law* treatments only (since norm question is not asked in the *Baseline*). Reference category: *Law* treatment. *** p < .001, ** p < .01, * p < .05.

V. Conclusion

Using a standard paradigm from experimental economics, this paper shows that customary law has a strong beneficial effect. It helps experimental participants overcome a social dilemma. If there are no sanctions, the effect basically coincides with the behavioural effect of what public international law calls comity. In essence, customary law governs behaviour since normative expectations and behaviour coevolve. Whether the rule invokes the authority of the law is at best immaterial, if not detrimental. Yet the authority of the law becomes instrumental as soon as there are sanctions. If combined with comity, sanctions crowd out some of the beneficial effect. If the rule originates in law, however, the authority of the law and the threat of sanctions reinforce each other.

There is an obvious gap between the behaviour of students in a computer lab and state practice. It could well be that states are not affected by normativity the same way as individuals. But as laid out in section 2, the debate in public international law does not rest on factors that are specific to states being the actors. The underlying claims address actors in general, and should therefore also be valid for individuals. It is these claims we address.

Experimental law and economics is still a very young discipline. Legal readers often expect a one to one mapping between the legal issue and the experiment. Experiments

cannot but disappoint those readers. Of necessity, any experiment is contextually poorer than the real life phenomenon it aims to understand. The purpose of an experiment is not being realistic but to isolate specific factors the theoretical debate believes to be important, and determine their causal effect. Experiments solve the identification problem that plagues the analysis of field data. Through random assignment, in an experiment the observed difference in behaviour between two treatments can only be caused by the deliberate difference in the design. One is sure in which direction the arrow of causality points, and one may safely conclude that the effect is not caused by unobserved, omitted variables. Experimental findings can never decide a complex legal issue. Experiments serve a different function but one that may be highly instrumental for law. Not so rarely, within a much richer legal discourse, one argument carries special weight. If this argument is empirical, experiments may make a valuable contribution. They may corroborate, or refute, this one argument. Such has been the purpose of our study. As we have laid out in the introduction and in the section on the legal debate, fairly abstract claims have been brought forward to invalidate the normative force of customary law in international relations. We have taken these arguments at face value, and refuted them.

An obvious next step for studying the coevolution of behaviour and normative expectations in public international law would be to qualitatively explore normativity in international relations (cf. Risse 1999**;** Risse 2000). In one respect, the character of states as corporate actors makes this even easier than with individuals. While the formation of the individual will occurs in her *forum internum*, the formation of the corporate will is open to public scrutiny. Through freedom of information legislation, even internal government deliberation is made accessible. While the proof may not be provided in this paper, it seems plausible that normativity in international relations is not fundamentally different from normativity in personal relations. Normativity is a key ingredient of human sociality (Wyman, Rakoczy et al. 2009). It might be scaled up to international law but not a substantially different force. If that could be shown to be true, the experimental findings reported in this paper would not only help international lawyers repel premature criticism of customary law. Our findings would even provide international lawyers with a conceptual framework for understanding one of the oldest sources of law: custom.

References

ANZILOTTI, DIONISIO (1955). Corso di Diritto Internazionale. Padova,, Cedam.

ARLEN, JENNIFER, MATTHEW L. SPITZER, et al. (2002). "Endowment Effects within Corporate Agency Relationships." Journal of Legal Studies **31**: 1-37.

AUMANN, ROBERT J. and LLOYD S. SHAPLEY (1994). Long Term Competition - A Game Theoretic Analysis. Collected Papers I. R. J. Aumann. Cambridge, MIT Press**:** 395-409.

BERNASCONI, MICHELE, LUCA CORAZZINI, et al. (2010). 'Expressive'Obligations in Public Good Games. Crowding-in and Crowding-out Effects. Working Papers.

BOHNET, IRIS and ROBERT D. COOTER (2001). Expressive Law. Framing or Equilibrium Selection?

BOLTON, GARY E. and ELENA KATOK (1998). "An Experimental Test of the Crowding Out Hypothesis. The Nature of Beneficent Behavior." Journal of Economic Behavior & Organization **37**: 315-333.

BORGES, GEORG and BERND IRLENBUSCH (2007). "Fairness Crowded Out by Law. An Experimental Study of Withdrawal Rights." Journal of Institutional and Theoretical Economics **163**: 84-101.

CHAN, KENNETH S., ROB GODBY, et al. (2002). "Crowding-out Voluntary Contributions to Public Goods." Journal of Economic Behavior & Organization **48**(3): 305-317.

CHAUDHURI, ANANISH (2011). "Sustaining Cooperation in Laboratory Public Goods Experiments. A Selective Survey of the Literature." Experimental Economics **14**: 47-83.

CHINEN, MARK A. (2001). "Game Theory and Customary International Law. A Response to Professors Goldsmith and Posner." Michigan Law Review **23**: 143-189.

CROSON, RACHEL T.A. and JASON SCOTT JOHNSTON (2000). "Experimental Results on Bargaining Under Alternative Property Rights Regimes." Journal of Law, Economics and Organization **16**: 50-73.

ECKEL, CATHERINE C., PHILIP J. GROSSMAN, et al. (2005). "An Experimental Test of the Crowding Out Hypothesis." Journal of Public Economics **89**: 1543-1560.

ELIAS, OLUFEMI (1995). "The Nature of the Subjective Element in Customary International Law." International and Comparative Law Quarterly **44**: 501-520.

ENGEL, CHRISTOPH (2008). "Learning the Law." Journal of Institutional Economics **4**: 275-297.

ENGEL, CHRISTOPH (2009). "Erga omnes. Why does Public International Law Ignore Privity of Contract?" Journal of Institutional and Theoretical Economics **165**: 24-28.

FEHR, ERNST and SIMON GÄCHTER (2000). "Cooperation and Punishment in Public Goods Experiments." American Economic Review **90**: 980-994.

FEHR, ERNST and SIMON GÄCHTER (2001). Do Incentive Contracts Crowd Out Voluntary Cooperation?

FEHR, ERNST and SIMON GÄCHTER (2002). "Altruistic Punishment in Humans." Nature **415**: 137-140.

FELDMAN, YUVAL and JANICE NADLER (2006). "The Law and Norms of File Sharing." San Diego Law Review **43**: 577-618.

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

FREY, BRUNO and RETO JEGEN (2001). "Motivation Crowding Theory. A Survey of Empirical Evidence." Journal of Economic Surveys **15**: 589-611.

FUNK, PATRICIA (2007). "Is there an Expressive Function of Law? An Empirical Analysis of Voting Laws with Symbolic Fines." American Law and Economics Review **9**(1): 135-159.

GALBIATI, ROBERTO and PIETRO VERTOVA (2008). "Law and Behaviours in Social Dilemmas. Testing the Effect of Obligations on Commitment." Games and Economic Behavior **64**: 146-170.

GEISINGER, ALEX (2002). "A Belief-Change Theory of Expressive Law." Iowa Law Review **88**(1): 35-73.

GOLDSMITH, JACK and ERIC A. POSNER (1999). "A Theory of Customary International Law." University of Chicago Law Review **66**: 1113-1177.

GRECHENIG, KRISTOFFEL, ANDREAS NICKLISCH, et al. (2010). "Punishment Despite Reasonable Doubt. A Public Goods Experiment with Sanctions Under Uncertainty." Journal of Empirical Legal Studies **7**(4): 847-867.

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

GUZMAN, ANDREW T. (2008). How International Law Works. A Rational Choice Theory, Oxford University Press, USA.

HECKATHORN, DOUGLAS D. (1989). "Collective Action and the Second-Order Free-Rider Problem." Rationality and Society **1**: 78-100.

JANSSEN, MAARTEN C.W. and EWA MENDYS (2004). "The Price of a Price. On the Crowding Out of Social Norms." Journal of Economic Behavior & Organization **55**: 377-395.

KAHNEMAN, DANIEL, ILANA RITOV, et al. (1999). "Economic Preferences or Attitude Expressions? An Analysis of Dollar Responses to Public Issues." Journal of Risk and Uncertainty **19**: 203-235.

KELSEN, HANS (1952). Principles of International Law. New York, Rinehart.

LEDYARD, JOHN O. (1995). Public Goods. A Survey of Experimental Research. The Handbook of Experimental Economics. J. H. Kagel and A. E. Roth. Princeton, NJ, Princeton University Press**:** 111-194.

LOEWENSTEIN, GEORGE and DON A. MOORE (2004). "When Ignorance Is Bliss. Information Exchange and Inefficiency in Bargaining." Journal of Legal Studies **33**: 37-58.

MCADAMS, RICHARD H. (2000). "An Attitudinal Theory of Expressive Law." Oregon Law Review **79**: 339-390.

MCADAMS, RICHARD H. (2000). "A Focal Point Theory of Expressive Law." Virginia Law Review **86**: 1649-1729.

MCADAMS, RICHARD H. and JANICE NADLER (2005). "Testing the Focal Point Theory of Legal Compliance. The Effect of Third-Party Expression in an Experimental Hawk/Dove Game." Journal of Empirical Legal Studies **2**: 87-123.

MCADAMS, RICHARD H. and JANICE NADLER (2008). "Coordinating in the Shadow of the Law: Two Contextualized Tests of the Focal Point Theory aof Legal Compliance." Law and Society Review **42**: 865-898.

MENDELSON, MAURICE H. (1998). "The Formation of Customary International Law." Recueil des Cours **272**: 155-410.

MORELLI, GAETANO (1967). Nozioni di Diritto Internazionale. Padova,, CEDAM.

NADLER, JANICE and SHARI SEIDMAN DIAMOND (2008). "Eminent Domain and the Psychology of Property Rights: Proposed Use, Subjective Attachment, and Taker Identity." Journal of Empirical Legal Studies **5**: 713-749.

NIKIFORAKIS, NIKOS S. and HANS-THEO NORMANN (2008). "A Comparative Statics Analysis of Punishment in Public Good Experiments." Experimental Economics **11**: 358-369.

NORMAN, GEORGE and JOEL P. TRACHTMAN (2005). "The Customary International Law Game." American Journal of International Law **99**: 541-580.

NYBORG, KARINE and MARI REGE (2001). Does Public Policy Crowd Out Private Contributions to Public Goods?

POSNER, ERIC A. (2009). "Erga Omnes Norms, Institutionalization, and Constitutionalism in International Law." Journal of Institutional and Theoretical Economics **165**: 5-23.

RISSE, THOMAS (1999). "International Norms and Domestic Change. Arguing and Communicative Behaviour in the Human Rights Area." Politics & Society **27**: 529-559.

RISSE, THOMAS (2000). ""Let's Argue!". Communicative Action in World Politics." International Organization **54**: 1-39.

ROSENTHAL, ROBERT W. (1981). "Games of Perfect Information, Predatory Pricing and the Chain Store Paradox." Journal of Economic Theory **25**: 92-100.

SELTEN, REINHARD (1978). "The Chain Store Paradox." Theory and Decision **9**: 127-159.

STERN, BRIGITTE (2001). "Custom at the Heart of International Law." Duke Journal of Comparative and International Law **11**: 89-108.

TRIEPEL, HEINRICH (1899). Völkerrecht und Landesrecht. Leipzig,, C.L. Hirschfeld.

TVERSKY, AMOS and DANIEL KAHNEMAN (1981). "The Framing of Decisions and the Psychology of Choice." Science **211**: 453-458.

TVERSKY, AMOS and ITAMAR SIMONSON (1993). "Context-Dependent Preferences." Management Science **39**: 117-185.

TYRAN, JEAN-ROBERT and LARS P. FELD (2006). "Achieving Compliance when Legal Sanctions are Non-Deterrent." Scandinavian Journal of Economics **108**: 135-156.

WYMAN, EMILY, HANNES RAKOCZY, et al. (2009). "Normativity and Context in Young Children's Pretend Play." Cognitive Development **24**: 146-155.

YAMAGISHI, TOSHIO (1986). "The Provision of a Sanctioning System as a Public Good." Journal of Personality and Social Psychology **51**: 110-116.

ZEILER, KATHRYN (2010). "Cautions on the Use of Economics Experiments in Legal Analysis." Journal of Institutional and Theoretical Economics **166**: 178-193.

ZELMER, JENNIFER (2003). "Linear Public Goods. A Meta-Analysis." Experimental Economics **6**: 299-310.

Appendix: Instructions

In the following, we present the experimental instructions as well as the control questions for Treatment *Law-S*. All other treatments are reduced forms of *Law-S*:

- Omitting the box "Customary Law:…" we obtain *Comity-S*.
- Additionally omitting the shaded areas we obtain *Baseline-S*.
- Finally, by omitting Stage 3 of the instructions and Question 5 of the questionnaire, we obtain the 3 treatments without sanctions: *Law*, *Comity*, and *Baseline*.

General instructions for the participants

Welcome to our experiment!

If you read the following explanations carefully, you will be able to earn a substantial sum of money, depending on the decisions you make. It is therefore crucial that you read these explanations carefully.

During the experiment there shall be absolutely no communication between participants. Any violation of this rule means you will be excluded from the experiment and from any payments. If you have any questions, please raise your hand. We will then come over to you.

During the experiment we will not calculate in euro, but instead in taler. Your total income is therefore initially calculated in taler. The total number of taler you accumulate in the course of the experiment will be transferred into euro at the end, at a rate of

1 Euro = 60 Taler

At the end you will receive from us the **cash** sum, in euro, based on the number of taler you have earned.

The experiment consists of **30 periods**, and each period consists of **3 stages**. Participants are randomly divided into groups of four. Apart from yourself, your group therefore has 3 further members. During these 30 periods, the constellation of your group of four remains unchanged. **Hence, you are with the same people in the same group for 30 periods.** At the beginning, each group member is allocated a random number between 1 and 4. This number remains unchanged for the entire 30 periods.

Stage 1:

At the beginning of each period, each participant is given **20 taler** to work with, referred to henceforth as **endowment**. Your task is to decide upon how to use your endowment. You must decide how many of the 20 taler you wish to pay into a common **project**, and how many you wish to keep for yourself. The consequences of this decision are explained in more detail below.

Your **endowment** hence consists of **20 taler in each period**. You make a decision on your payments by typing whole numbers between 0 and 20 in the input field on your screen. Once you have keyed in your amount, press **Continue**. As soon as you have done this, you may no longer reverse your decision for this period.

Once all group members have made their decisions, you are told how much each individual group member has contributed to the project.

Your total income (in taler) therefore consists of two parts: (1) the taler income from the common project and (2) the taler you have retained.

Total income (in taler)	=	Income from the common project	+	Taler retained

The **income from the common project** is calculated as the total sum of all contributions to the project (within your group of four) times 0.4.

Income from the common project	=	total sum of all contributions to the project (within your group of four)	×	0.4

Example:
If the sum of contributions from all group members to the common project is 60 taler, you and each other group member receive an income from the project of $0.4 \times 60 = 24$ taler. If the group members have contributed a total of 9 taler to the project, you and each other group member receive a taler income from the project of $0.4 \times 9 = 3.6$.

If you contribute one taler from your endowment to the group project, the sum of contributions to the common project increases by 1 taler, and your income from the project increases by $0.4 \times 1 = 0.4$ taler. However, this also means that each individual other group member's income increases by 0.4 taler, so that the total income of the group increases by $0.4 \times 4 = 1.6$ taler. The other group members therefore also earn something from your contribution to the project. On the other hand, you profit from the contributions made by the other group members. For each taler contributed to the project by another

group member, you earn 0.4×1 = 0.4 taler. Hence, if each member of your group of four contributes 1 taler to the project, each of you receives 0.4×1×4=1.6 taler as income from the project.

Stage 2

In Stage 2, you will see a screen requesting you to answer the following questions:

1. Do you believe that there is a general **norm** in your group on an **appropriate minimum contribution** to the project? (Yes/No)

2. If yes, how high can this minimum contribution be expected to be? (Number between 0 and 20).

Customary Law:

For new law to originate, it is not necessary that the legislator pass a statute, or that the parties agree on an explicit contract. Customary law is equally valid and binding. Customary law comes into being if the large majority of those affected for a sufficiently long period behave in a sufficiently similar way. The fact that some contribute even more to the joint project does not prevent that a norm of customary law originates. Hence a rule of customary law may prescribe a minimum standard. Customary law may originate here in the lab as well.

From the second period onwards, you will receive information on the behavior of individual group members in past periods. In order to receive this, you will have to click on an appropriate **button** on your screen. This can be done as often as you like.

- Button "**contributions**": how much have the individual group members contributed to the common project?

Stage 3

In Stage 3 you learn how much the other group members have contributed to the common project in this period. You then have the possibility of reducing the other players' income by **distributing points**. Each point you distribute to the other players costs you 1 Taler and reduces the other players' income also by 1 Taler.

Do you have any further questions?

Control Questionnaire

1. Each group member has an endowment of 20 taler. Nobody (including you) contributes any taler to the project. What is:
   a. Your income from the common project? ……
   b. Your total income?……

2. Each group member has an endowment of 20 taler. You contribute 20 taler to the project. All other group members contribute 20 taler each to the project. What is:
   a. Your income from the common project? ……
   b. Your total income?……

3. Each group member has an endowment of 20 taler. You contribute 0 taler to the project. The other three group members contribute together a total of 30 taler to the project. What is:
   a. Your income from the common project? ……
   b. Your total income?……

4. Each group member has an endowment of 20 taler. You contribute 15 taler to the project. The other three group members contribute together a total of 5 taler to the project. What is:
   a. Your income from the common project? ……
   b. Your total income?……

5. After Stage 1 you have a total income of 30. Then you distribute 2 points to group member 1 and 3 points to group member 2. You also receive from the members of your group a total of 4 points. What is your total income now?

Chapter 3

Can we manage first impressions in cooperation problems?
An experimental study on "Broken (and Fixed) Windows"

Christoph Engel, Sebastian Kube, and Michael Kurschilgen

Abstract

Cooperation problems are at the heart of many everyday situations. In this paper, we propose a very simple and light-handed mechanism to sustain cooperation and test its performance in a rich laboratory environment. The mechanism moderates cooperation by controlling experiences, more specifically, it "manipulates" subjects' initial beliefs by providing them with selective information about (un)cooperative behavior in other, unrelated, groups. We observe that contributions are considerably sensitive to such selective information. First impressions participants happen to make predict subsequent behavior. Our results, however, suggest an asymmetry in the strength of the reaction – which might pose a limit on the effectiveness of the mechanism in natural settings.

Keywords: Public Good, Behavioral Uncertainty, Conditional Cooperation, Information, First Impressions, Broken Windows, Experiment

JEL: *C91, D03, D83, H41, K14, K42*

I. Introduction

Cooperation problems are at the heart of many everyday situations. For example, when it comes to protecting the environment, defending one's country, generating new knowledge, joining a political party, extending the infrastructure, or exploiting the opposite market side, agents face a social dilemma. Jointly they are best off if everyone contributes her fair share. But individually, free-riding on others' efforts yields the highest payoff. A large number of theoretical and empirical papers have explored how people should be expected to and how they actually do behave in such situations. Many empirical contributions make use of laboratory experiments where subjects participate in a prisoners' dilemma or in a public-good game. Absent institutional interventions like punishment (Fehr and Gächter 2000), contributions are heterogeneous in the beginning, but average cooperation quickly declines and most participants free-ride in the end (e.g. Andreoni 1988). These results have been stress-tested extensively, e.g., with respect to anonymity (Andreoni and Petrie 2004), culture (Herrmann, Thöni et al. 2008), group size (Isaac, Walker et al. 1994), efficiency (Glöckner, Irlenbusch et al. 2010), or framing (Goerg and Walkowitz 2010). Interventions that mitigate the dilemma are not easily designed. Effective interventions tend to be heavy-handed, often altering the incentive structure of the game such that free-riding is not in an individual's self-interest any more (e.g. Ostrom, Walker et al. 1992; Falkinger, Fehr et al. 2000; Fehr and Gächter 2002; Gürerk, Irlenbusch et al. 2006; Glöckner, Irlenbusch et al. 2010). In this paper, we propose a very simple and light-handed mechanism and test its effectiveness by using laboratory experiments.

Our mechanism is based on an observation that has previously been made in public-goods experiments, namely that (initial) group composition and initial cooperation rates significantly affect the future development of cooperation rates in a particular group (Burlando and Guala 2005; Gächter and Thöni 2007; Beckenkamp, Engel et al. 2009). A likely reason for this behavioral pattern is provided by the influential concept of reciprocity and conditional cooperation (Fischbacher, Gächter et al. 2001; Dufwenberg and Kirchsteiger 2004; Falk and Fischbacher 2006; Fischbacher and Gächter 2010, among many others). If persons are sensitive to the behavior of other group members, those groups who cooperate little in the beginning will become even less cooperative over time, while groups with substantial cooperation in the beginning are able to sustain cooperation over time. This suggests that outsiders might moderate cooperation by controlling experiences. Our mechanism is even less invasive in that it

is confined to the first impressions subjects happen to make. More precisely, the idea is to "manipulate" initial beliefs by providing participants with selective information about (un)cooperative behavior in other, unrelated, groups. The information affects initial beliefs about others' behavior, which in turn changes how I myself behave initially. A virtuous (or vicious, depending on the kind of information that is provided) cycle starts: via beliefs, our manipulation determines cooperativeness in the beginning, which in turn determines cooperativeness later on.

Our intervention idea is in the spirit of James Q. Wilson's Broken Windows Theory (Wilson and Kelling 1982) which is based on the work of American sociologist Philip Zimbardo. In 1969, Zimbardo abandoned two identical cars in two different locations: the Bronx, NYC and Palo Alto, California. "The license plates of both cars were removed and the hoods opened to provide the necessary releaser signals" (Zimbardo 1969). In the Bronx, the abandoned car was stripped and demolished after only 26 hours, the result of 23 separate incidents of vandalism. In contrast, the car in Palo Alto still sat unmolested after the course of an entire week. Zimbardo then decided to provide an example of vandalism to the affluent and seemingly non-violent neighborhood of Palo Alto. So he and two graduate students of his took a sledgehammer and started bashing the car. After they had taken the first blow, observers shouted encouragement and finally joined in the vandalism, until the car was completely wrecked (Zimbardo and Ebbesen 1969).

In order to clearly identify the short- and long-run effects of our mechanism we run laboratory experiments, which have the benefit of providing a sufficient degree of control over the environment. The environment that we use has been prominent lately to study social dilemmas and cooperation problems. It is a complex public-good game (more precisely, a voluntary contribution mechanism) with decentralized sanctions (Fehr and Gächter 2000) and counter-punishment opportunities (Nikiforakis 2008). We have chosen this rich environment, since it mimics several potentially important features available in natural environments. Moreover, it is sufficiently complex, so that conflicting behavioral norms might emerge – which allows us to test the effect of both favorable and unfavorable first impressions.

In the baseline, subjects play the game for 10 periods without any exogenous interventions. In two treatments (FAV and UNFAV), prior to making their first contribution decision in the game, we provide subjects with selected data from the baseline. Specifically, we show them graphs of the development of mean

contributions over time from selected cooperative (resp. uncooperative) groups. The information that we give is unfavorable for cooperation in treatment UNFAV – similar to Zimbardo, we "attack the car with a hammer" and explore whether "the car will then go to ruins", i.e., whether cooperation breaks down in the public-good experiment. Contrarily, in treatment FAV we reverse the manipulation by exposing subjects to favorable initial impressions – i.e., similar to Wilson, metaphorically speaking we repair the windows in a bad neighborhood, and investigate whether newly arriving inhabitants behave more cooperatively.

We find that cooperation rates strongly react to the treatment manipulation.[1] Unfavorable information substantially reduces initial contributions to the public good. Moreover, the effect is long-lasting and translates into future development of groups' cooperation behavior. Contributions in treatment UNFAV do not recover during the experiment. More importantly even for policy making, the favorable selective information in treatment FAV prevents cooperation rates from decaying over time. Taken together, our results suggest that subjects' behavior is indeed conditional on initial expectations of how others are likely to behave, and that these expectations are open to purposeful intervention.

The next section embeds our study into the existing literature. Subsequently, we explain the design of the experiment. Section 4 presents and discusses the experimental results of our main study, and features an additional experiment that explores whether the effect of our manipulation is mediated by subjects' initial expectations. Section 5 concludes the paper.

II. Related Literature

Our study closely relates to the experimental literature on reciprocity and conditional cooperation. This literature suggests that a substantial fraction of subjects usually consists of conditional cooperators, i.e., people who are willing to cooperate, provided a sufficiently large fraction of the population does the same. (Fischbacher, Gächter et al. 2001; Mengel 2007; Fischbacher and Gächter 2010). For conditional cooperators, information about contribution patterns in other groups has a first- and a

[1]As we demonstrate in a second series of experiments (reported below in Section 4), the effect of our manipulation is mediated via subjects' initial beliefs. Beliefs about others' initial contribution decisions are much more pessimistic when unfavorable information is provided.

second-order effect. Given the uncertainty about the composition of the group of which they happen to be a member, information about other groups helps them form beliefs. Moreover, each participant knows that each other participant has received the same information. This makes it possible to also form a second-order belief, based on knowing that the remaining group members have received the same information. The finding that, in our experiment, cooperation rates are sensitive to the initial information-manipulation is fully reconcilable with the idea that a non-negligible fraction of the population acts as if they were conditional cooperators. This yields further support for the existence and importance of conditional cooperation.

The research question covered here is also closely related to the literature on institutional design, specifically the branch that explores interventions aiming at raising contributions to public goods. Usually, the proposed mechanisms alter the incentive structure such that free-riding becomes less attractive, or even a dominated strategy for self-centered money maximizers. For example, in (Falkinger, Fehr et al. 2000) the payoff structure is changed such that "each individual gets a reward or has to pay a penalty depending on the deviation of its contribution from the mean contribution." In other studies, group composition is changed such that the payoff structure is changed implicitly. For example, in (Gunnthorsdotir, Houser et al. 2007), subjects are re-matched every period according to their cooperativeness in the previous round; which is found to raise cooperation levels. Likewise, if groups have a chance to exclude free-riders, this improves cooperation in a dilemma setting (Cinyabuguma, Page et al. 2005; Croson, Fatas et al. 2008), as does a mechanism that allows members to self-select into groups (Page, Putterman et al. 2005), in particular if free-riders are effectively excluded by a rule that sacrifices a portion of the group income to outsiders (the Red Cross, as it was , Brekke, Hauge et al. 2009). Our study differs from this literature in that we leave the incentive structure of the game completely unchanged. All we alter are the first impressions participants happen to make – which is an option that should naturally be available in any public good game, and in many real-life social dilemmas.

Parts of the legal literature, particularly those at the intersection of law and economics, have also been asking how socially desirable behavior and/or compliance with the law can be brought about. Again, a prominent approach is to change the incentive structure, e.g., by increasing the expected costs of breaking the law. For example, Braga, Weisburd et al. (1999) report evidence from a field experiment that randomly exposed 12 of 24 matched violent crime places in Jersey City to intense

police scrutiny and intervention. In the places chosen, crime rates dropped substantially, while they did not in the unaffected places. In a similar vein, in a series of sociological field experiments, when there were signs of disorder, like graffiti, abandoned shopping carts, littering or bicycles locked where they were not supposed to be, this induced passers-by to also break these and other rules (Cialdini, Reno et al. 1990; Keizer, Lindenberg et al. 2008; Ramos and Torgler 2010). In laboratory experiments, Galbiati and Vertova (2008) demonstrate that cooperation behavior increases when an explicit expectation is spelt out and enforced by (non-deterrent) sanctions (see also Kube and Traxler 2010).

Besides, our findings also underline the power of information, as it has also been observed in other contexts. For example, previous work on voting behavior points out that information in the form of polls (Forsythe, Myerson et al. 1993; Forsythe, Rietz et al. 1996; Klor and Winter 2007) or of cheap-talk electoral campaigns (Corazzini, Kube et al. 2010) affect subsequent voting outcomes. Similarly, information gathered during pre-play communication[2] strongly affects subjects' decision in subsequent coordination games (Blume and Ortmann 2007), as does information from preceding asset-market outcomes (Kogan, Kwasnica et al. 2010) or information about group members' previous decisions (Weber 2006). Interestingly, the literature on coordination games comes to the conclusion that pre-play information promotes Nash-equilibrium play (assuming self-centered money maximizers). By contrast, our experimental results show that the opposite might happen in a cooperation game like the one reported here, where the socially efficient outcome is usually not part of the subgame-perfect Nash equilibrium. Pre-play information may indeed promote socially efficient outcomes.

III. Experimental Design, Procedure, and Behavioral Predictions

Experimental Design: The baseline treatment is a typical public-good game with punishment and counter-punishment opportunities as implemented by Nikiforakis (2008).[3] The basic game features n=4 players and consists of three stages:

[2] See alsoCrawford (1998) for a general survey of experiments on communication via cheap talk.

[3]We have chosen this design because it provides a rich environment in which different behavioral norms may emerge. Moreover, this design has the advantage to leave sufficient room for testing the effect of both favorable and unfavorable first impressions. Had we tested a mere voluntary contribution mechanism, previous experiments would have indicated that cooperation is very difficult to sustain in the first place. Conversely, had we chosen a public good game with one-step punishment, successful

At the beginning of stage 1 ("contribution stage"), players are endowed with 20 tokens each. Players then decide simultaneously and independently how much of their endowment they want to contribute to a public account. We denote this decision with c_i. Each token that is contributed to the public account increases the payoff of each player in the group by 0.4 tokens (i.e., the MPCR is α=.4). Each token unspent increases a player's own payoff by one token. The preliminary payoff at the end of stage 1 is thus given by.

$$\pi_i^1 = 20 - c_i + 0.4\sum_{h=1}^{4} c_h$$

Table 1: Punishment points p_{ij} per player j and costs C(p_{ij}) for punisher i

p_{ij}	0	1	2	3	4	5	6	7	8	9	10
$C(p_{ij})$	0	1	2	4	6	9	12	16	20	25	30

At the beginning of stage 2 ("punishment stage"), players are informed about every group member's contribution to the public account. Every player i then has the opportunity to reduce the income of each other group member j by assigning costly punishment points p_{ij}. Each punishment point received reduces a player's income from stage 1 by ten percent. At the same time, each punishment point assigned reduces one's own payoff according to the cost function given in Table 1. The preliminary payoff at the end of stage 2 is thus given by:

$$\pi_i^2 = \pi_i^1 \frac{1}{10}\max\left\{0, 10 - \sum_{j\neq i} p_{ji}\right\} - \sum_{j\neq i} C(p_{ij}) .$$

At the beginning of stage 3 ("counter-punishment stage"), players observe who punished them by how much in stage 2. They then have the opportunity to counter-punish the punishers by assigning them counter-punishment points cp_{ij}. The punishment technology is the same as in stage 2. Each counter-punishment point received reduces a player's preliminary income from stage 2 by ten percent and each

cooperation would have been very likely. By contrast, according to the existing literature, if we also add the counter-punishment option, expectations are in the middle between both extremes.

counter-punishment point assigned reduces one's own payoff as given in Table 1, but the cost of counter-punishment also depends on this player's punishment decision.[4] The final payoff at the end of stage 3 is thus given by:

$$\pi_i^3 = \pi_i^2 \frac{1}{10} \max\left\{0, 10 - \sum_{j \neq i} cp_{ji}\right\} - \sum_{j \neq i} [C(p_{ij} + cp_{ij}) - C(p_{ij})]$$

In treatments FAV and UNFAV, the underlying game structure remains unchanged. Compared with the baseline the only change is that we append the experimental instructions by one additional page. The page contains selected information about unrelated groups who previously played this game – and players know that the information is actually taken from previous experiments. They are also told that the information is selective and that all players in their group receive the same information. The information we give is the development of mean contributions over time from four selected groups. Furthermore, we provide them with selected data about the number of persons choosing to contribute zero in the first, resp. in the last period, and the average amount of counter-punishment that was meted out in these experiments. The way this data is presented and the selection of the four groups are such that they constitute a positive (i.e. cooperative) impression in treatment FAV, while they constitute a negative (i.e. uncooperative) impression in treatment UNFAV.[5] Thus, the treatment manipulations allow us to study to what extent different contents of pre-play information affects behavior.

Procedure: The experiments were run in two different locations, namely in London (UK) and Bonn (Germany). The baseline treatment was played in both locations. As will be seen below, the cooperation rates in the baseline treatment differ between these two locations. The observed cooperation rates in this environment are rather low in London, while they are rather high in Bonn .[6] Taking advantage of this

[4]That is, if *i* assigns counter-punishment points to *j*, the specific costs of the counter-punishment points depend on the number of punishment points that *i* has assigned to *j* on stage two. For example, if *i* had already assigned four points to *j* on stage 2, assigning him a single counter-punishment point on stage three costs 3 tokens. Hence the costs of counter-punishment are given by $C(p_{ij}+cp_{ij})-C(p_{ij})$.

[5]The Appendix contains the exact wording of these two pages including the graphs, as well as a translation of the general instructions that were used in all treatments.

[6]The corresponding data that we use for our baseline treatments BASEFAV and BASEUNFAV are taken from Nikiforakis (2008) and Beckenkamp, Engel et al. (2009), respectively. The data reported from treatments FAV and UNFAV were collected by us. They are novel and not reported elsewhere. We are thankful to Nikos Nikiforakis for providing us with his data.

observed variation between the two locations, we run treatment FAV in London and treatment UNFAV in Bonn. Consequently we refer to the London baseline as BASEFAV and to the Bonn baseline as BASEUNFAV. Apart from the difference in location and thus the language of the instructions, the procedure was as similar as possible in all treatments.

All experiments were programmed using z-Tree (Fischbacher 2007). Subjects in Bonn were invited using ORSEE (Greiner 2004). The composition of subjects with respect to age, gender and field of studies was similar between locations and between treatments. Subjects were not allowed to participate in more than one treatment (between-subject design). When subjects arrived in the lab, they were seated in separate cubicles. The experimental instructions were then handed out to them and read out aloud. This was done to create common knowledge and to ensure that everybody had read and understood the instructions. Additionally, subjects had to take a quiz and could pose comprehension questions in private before the game started. The instructions were written in neutral language, avoiding potentially loaded terms like punishment or public good (cp. Appendix 1).

After subjects had finished reading the instructions, in treatments FAV and UNFAV they were provided with the additional information sheet (cp. Appendix 3 and 4). Subjects were then randomly divided into groups of four and played the above described game repeatedly for ten consecutive periods. A partner protocol was used, i.e., the group composition stayed constant over the entire 10 periods of anonymous interaction. In the end, subjects were privately paid their cumulated earnings and left.

We had 60 participants (3 sessions) in treatment FAV in the Laboratory for Experimental Economics at the University of Bonn in 2009. Participants received their accumulated earnings from the experiment (1 token = Euro 0.04) and an additional show-up fee of 5 Euro. On average, a session lasted 60 minutes and subjects earned Euro 14.46. The four sessions for treatment UNFAV were run in 2010 at the experimental laboratory of Royal Holloway, University of London, UK. We used the same lab and recruited 64 students from the same subject pool as in Nikiforakis (2008). We also used the same software. Again, participants received their accumulated earnings from the experiment (1 token = 0.04 British Pounds). On average, a session lasted 60 minutes and subjects earned 9.06 Pounds. In the two baseline treatments, analogous procedures had been used. The datasets comprise 48 subjects in BASEFAV and 68 participants in BASEUNFAV. Additional details on the

two baselines can be found in Nikiforakis (2008) and Beckenkamp, Engel et al. (2009), respectively.

Behavioral Predictions: For rational, self-centered money maximizing players, the game at hand is a cooperation problem. Using backwards induction, it is straightforward that the unique subgame perfect equilibrium is to contribute nothing to the public good at the first stage, and not to punish nor counter-punish at subsequent stages. This leads to an equilibrium payoff of 20 tokens per period and player. By contrast, the socially efficient outcome would be achieved if everyone contributed their entire endowment, in which case every player would earn 32 tokens per period. But in that case, each player would have an individual incentive to free-ride on the others' contributions, which would yield him a payoff of 44 tokens, ceteris paribus. Since we announce the number of periods, through unraveling this is also the prediction for the repeated game. The standard prediction of no cooperation not only holds true in the baseline treatment. Providing subjects with additional information about other groups does not change the equilibrium prediction. Therefore, with rational, self-centered money maximizing players we should not expect to observe different behavior in treatments FAV or UNFAV.

By contrast, things might change as soon as we allow for social preferences, in particular if we expect a substantial fraction of the population to act as conditional cooperators. If conditional cooperators are sufficiently optimistic about the cooperativeness of their interaction partners, cooperation gives them higher utility than defection. Facing a public good, conditionally cooperative subjects should therefore base their decision on what they believe other subjects to do. The beliefs are likely to be related to other group members' behavior in the previous period(s). Yet, upon the first encounter (in the first period), a conditional cooperator needs to form initial expectations of how others are likely to behave. This is where our treatment manipulation might make a difference. Pre-play information about other, unrelated groups might influence a participant in forming initial beliefs, which should then guide her first-period decision. If her (conditionally cooperative) group peers were also influenced by the selective pre-play information, then group contributions in the first period are likely to mirror the pre-play information. Her beliefs for the second period will thus be very similar again, and so will contributions be in the second period; and similarly in subsequent periods. In that case, we would expect to observe more cooperation in treatment FAV than in BASEFAV, and less cooperation in treatment UNFAV than in BASEUNFAV throughout the game.

IV. Results

In the following, we will first briefly describe the results from the two baseline treatments BASEFAV and BASEUNFAV. As will be seen, observed cooperation rates are low in London while they are high in Bonn. Afterwards, we move on to show that the treatment manipulations have a substantial and significant influence on subject's behavior. Compared to the respective baseline treatment, cooperation rates are lower in treatment UNFAV while they are higher in treatment FAV. As will be seen, this is likely to be driven by the change in first period's contributions which, in turn, are sensitive to our manipulation of first impressions. In line with this, group's behavior in the first period is strongly correlated with its future development, as we subsequently show. In the last subsection, additional experiments further underline our result by explicitly exploring how beliefs react to the treatment manipulations.

BASELINES:
Figure 1 shows the mean development of contributions over time in BASEFAV and BASEUNFAV. As can be seen, the environment in which subjects interact is complex and allows for different behavioral patterns.[7] Without any manipulation, cooperation rates are rather low in London (BASEFAV). The mean contribution in BASEFAV is 8.72. By contrast, subjects in Bonn (BASEUNFAV) display a high degree of cooperation. They contribute on average 16.45 to the public good. We therefore test in Bonn whether poor first impressions reduce the tendency to contribute much, and thereby trigger a vicious cycle. Conversely we test in London whether promising first impressions increase the willingness to give more initially, and thereby trigger a virtuous cycle.

[7]While this observation, in particular the behavioral differences between locations, is interesting *per se*, here we do not want to elaborate on this point but would like to refer the interested reader to (Beckenkamp, Engel et al. 2009) for an extensive discussion.

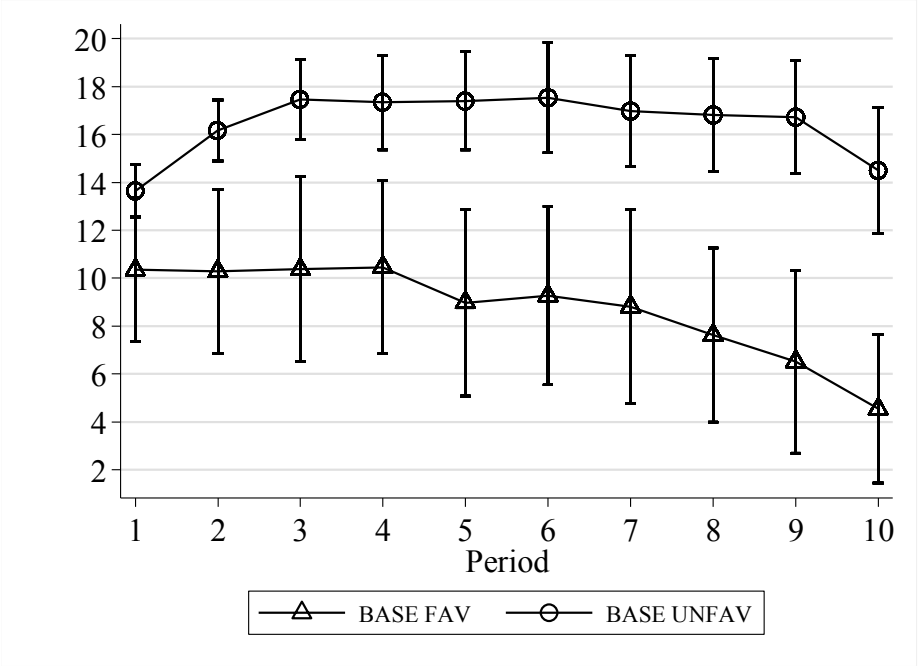


Figure 1: Mean Contribution over time in BASEFAV and BASEUNFAV
Notes: With 95% Confidence intervals around the period means. Group clustered standard errors. BASEFAV(Nikiforakis 2008) has 48 subjects (12 groups). BASEUNFAV(Beckenkamp, Engel et al. 2009) has 68 subjects (17 groups).

BREAKING WINDOWS:
What happens if we provide subjects with additional information that conveys a negative pre-play impression? As can be seen in Figure 2, this experimental manipulation has a substantial effect on contributions. After "breaking the window" with an "experimental hammer", behavior changes radically. Whereas in BASEUNFAV subjects choose full contribution in 60% (411/680) of all cases, in UNFAV they only do so in 20% (120/600) of all cases. Also at the opposite end of the contribution scale the effect is evident. In BASEUNFAV subjects only in 8.5% (58/680) of all cases contribute less than 5 tokens. With the treatment manipulation, this fraction grows to almost 32% (191/600). Overall, the mean contribution drops significantly from 16.45 in BASEUNFAV down to 10.71 in UNFAV (Mann Whitney ranksum test, N = 32, p=0.0024, two-sided). Hence, our "experimental hammer" indeed creates uncooperative behavior in an otherwise cooperative environment, just like

Zimbardo's sledgehammer generated Bronx-like behavior in Palo Alto. Moreover, the effect is long-lasting: the initial manipulation of expectations still leads to considerably lower cooperation even after ten periods of interaction, i.e. after 9 possible opportunities to revise expectations.

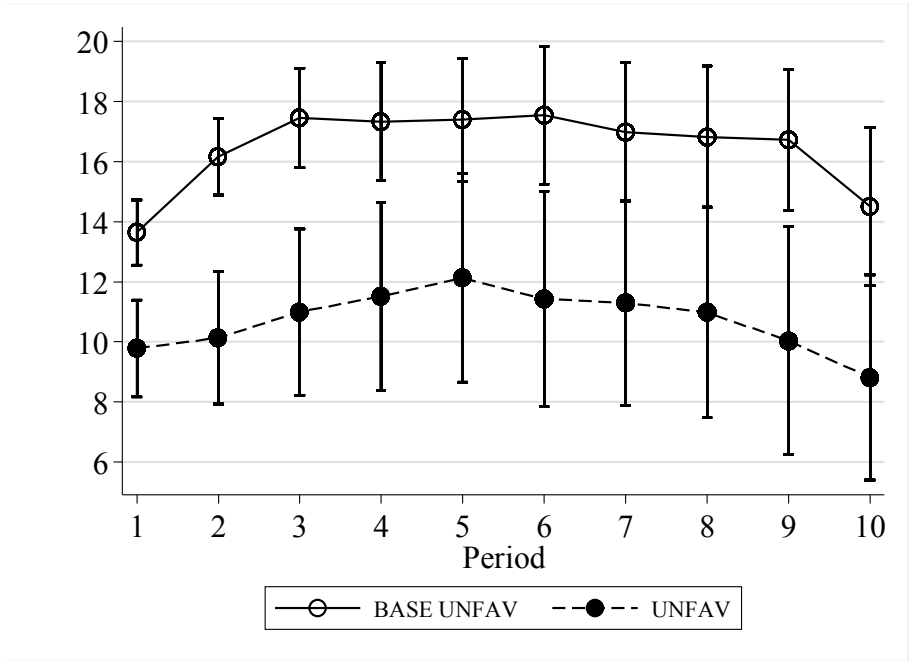


Figure 2: Mean Contribution over time in BASEUNFAV and UNFAV
Notes: With 95% Confidence intervals around the period means. Group clustered standard errors. BASEUNFAV (Beckenkamp, Engel et al. 2009) has 68 subjects (17 groups). UNFAV has 60 subjects (15 groups).

FIXING WINDOWS:
From a policy perspective, the opposite situation is even more interesting. Is the mere fact that one "repairs windows" enough to foster cooperation? To test this hypothesis, in treatment FAV we provide participants with favorable first impressions. Inspecting Figure 3, it is plain that repairing windows is less powerful than breaking them. Descriptively, once participants have seen graphs from groups that have cooperated successfully, they contribute more. Average contributions increase from 8.7 in BASEFAV to 12 in FAV. This increase of almost 40% is remarkable, though it falls short of being statistically significant (ranksum test, p=.1567, two-sided). Still, the behavioral pattern seems to be strongly influenced by the favorable manipulation.

While cooperation is sustained in FAV, contribution rates quickly decline in BASEFAV. Consequently, from the second period on, mean contributions in BASEFAV are outside the 95% confidence interval of mean contributions in FAV after participants have seen graphs from cooperative groups. Contributions rise sharply for the first three periods. They then stay more or less stable until period six. There is a slight decay for the final periods. Compare this to the data from treatment BASEFAV. Contributions never rise. They decay from the fourth period on. This difference of trends is significant (Mann Whitney over means of first differences, N = 28, p = .0243). Our data thus suggests that fixing windows pays. The beneficial effect only takes longer to unfold.

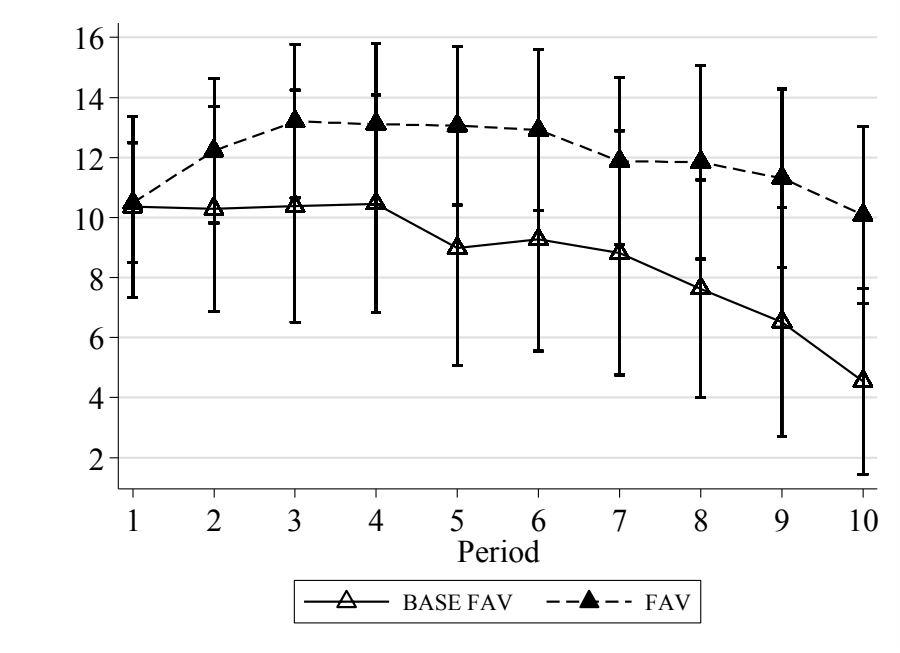


Figure 3: Mean Contribution over time in BASEFAV and FAV
Notes: With 95% Confidence intervals around the period means. Group clustered standard errors. BASEFAV(Nikiforakis 2008) has 48 subjects (12 groups). FAV has 64 subjects (16 groups).

INITIAL CONTRIBUTIONS AND FUTURE DEVELOPMENT:
How does this very simple and light-handed mechanism of providing subjects with information from unrelated groups create these strong differences in the overall

cooperation rate? If we take a closer look at our data, the answer to this question seems to be twofold.

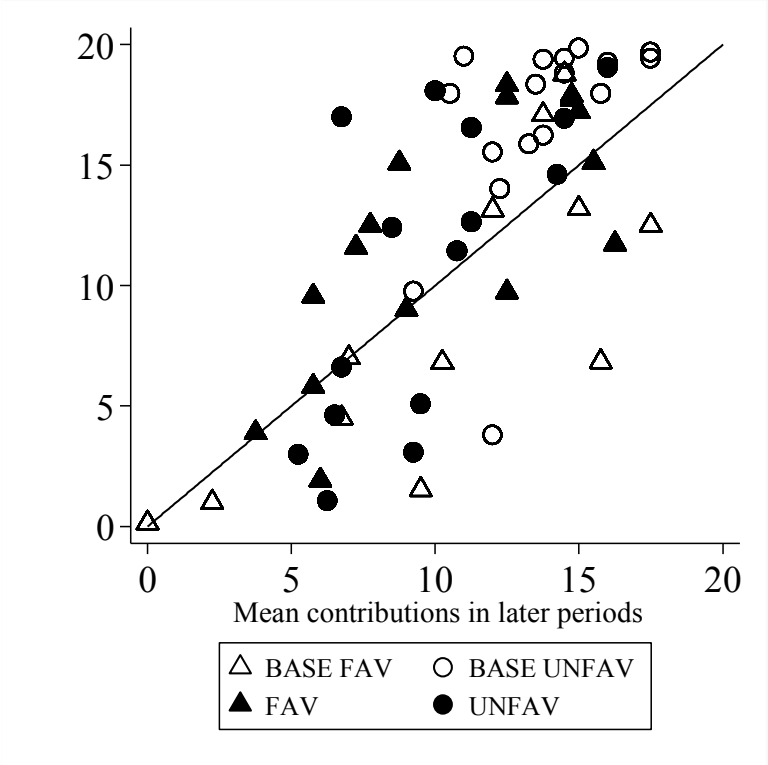


Figure 4: Initial contributions vs. future group development
Notes: Every entry in the graph represents one specific 4-person group. There are 12 groups in BASEFAV, 17 in BASEUNFAV, 16 in FAV and 17 in UNFAV. "Later periods" refers to the mean contribution in periods 2 to 10. The 45 degree line depicts all points in which the average contribution in period 1 is identical to the average contribution in later periods.

First, we find that in all treatments contribution behavior in period 1 significantly predetermines cooperativeness in later rounds. This can best be seen when we look at the development in individual groups. Figure 4 compares the average contribution to the public good in period 1 with the average contribution in periods 2 to 10. There is a clear correlation between initial and future cooperation. The correlation is strong and significant, over all treatments (Spearman's rho=.6981, N = 60, p<.001) as well as for each treatment separately.

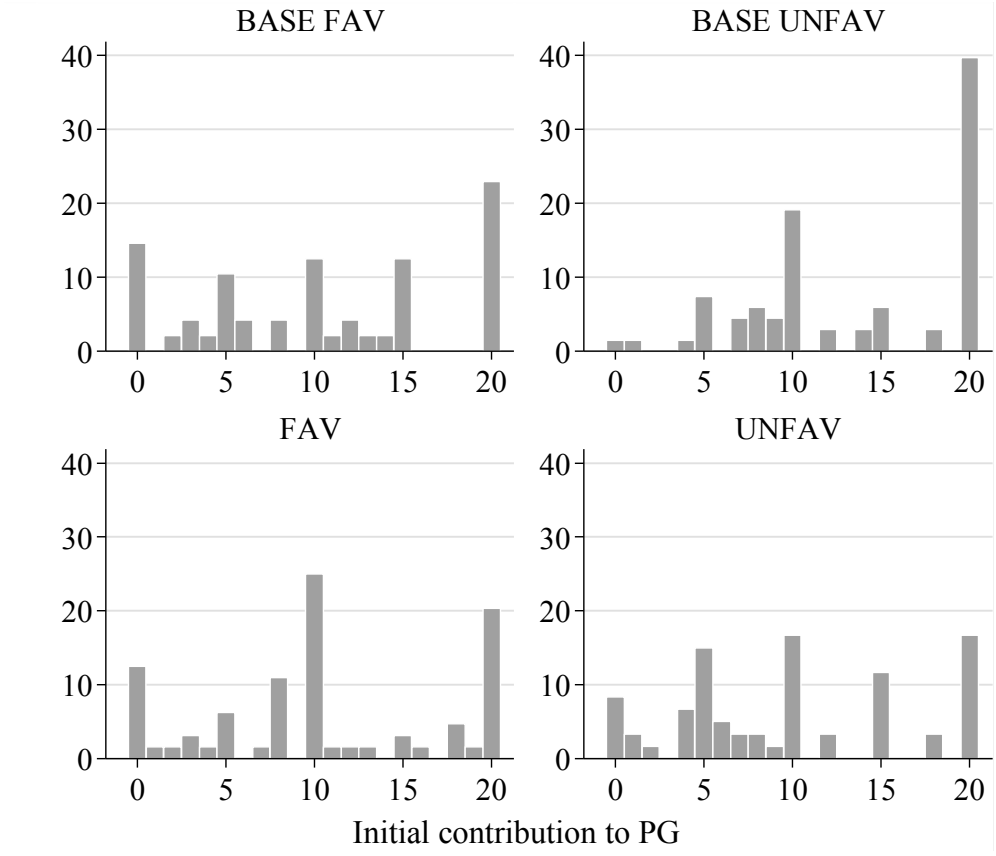


Figure 5: Initial contributions to the public good
Notes:BASEFAV has 48 subjects, BASEUNFAV 68, FAV 64, and UNFAV 60.

Second, we observe that our experimental manipulation has a substantial effect on initial contributions to the public good (cp. Figure 5). In BASEUNFAV, 40% of subjects go for full contribution in the first period. Yet if they receive the unfavorable information in treatment UNFAV, only 16% do so. The mean contribution in the first period drops by almost 30%, from 13.64 in BASEUNFAV down to 9.78 in UNFAV; the difference being significant (ranksum test, p=.0007, two-sided). Our favorable manipulation does not have the same power. Our data suggest that "fixing windows" in an uncooperative environment is not as straightforward as breaking them in an otherwise cooperative environment. The favorable information increases the mean

contribution in the first period only very slightly, from 10.35 in BASEFAV to 10.48 in FAV. This difference is not significant.

Taken together, these two findings provide a possible explanation for the overall difference in cooperation that we observe between treatments. Contributions are path-dependent, i.e., they depend on contributions in the first period. These first contributions, in turn, are apparently manipulable by our very simple mechanism (though the strength of this manipulation seems to be asymmetric). Thus, if people are provided with unfavorable information, i.e., if they observe "broken windows", they start with low cooperation rates and also stay at low cooperation rates in the long run; favorable information at least makes it easier to sustain cooperation in the long run. This suggests that subjects seem to form an initial expectation about how others are likely to behave and base their decision on this expectation. The initial expectations then appear as some kind of self-fulfilling prophecy: The mere pre-game expectation of an uncooperative environment leads to an irrevocable loss of cooperation throughout the game (and vice versa).

PRE-GAME EXPECTATIONS:

To shed more light on the role of expectations, we subsequently ran additional experiments in 2010 at the Econ Lab of the University of Jena, Germany. In these experiments, subjects received exactly the same instructions and had to answer the same control questions as in the games above. However, instead of actually playing the game they were then told that these experiments had been conducted before. Their task was to guess how much those previous participants contributed on average to the common project in the first period (rounded to the next integer). Moreover, subjects had to state how confident they were in this expectation (on a Likert scale ranging from 1 to 7, where 1 means "being very unsure" and 7 means "being very sure" about the belief).[8]

Altogether, 96 subjects participated in these experiments. In the first treatment, subjects received the instructions from the baselines. In the second treatment, they received the instructions of treatment FAV. Finally in the third treatment they received the instructions of treatment UNFAV.

[8] In the belief elicitation stage, subjects had to guess a whole number between 0 and 20. If the guess was exactly correct, they received 4 Euros. If the guess was wrong, they received 0 Euros. The confidence question was not incentivized. See Appendix 5 for details.

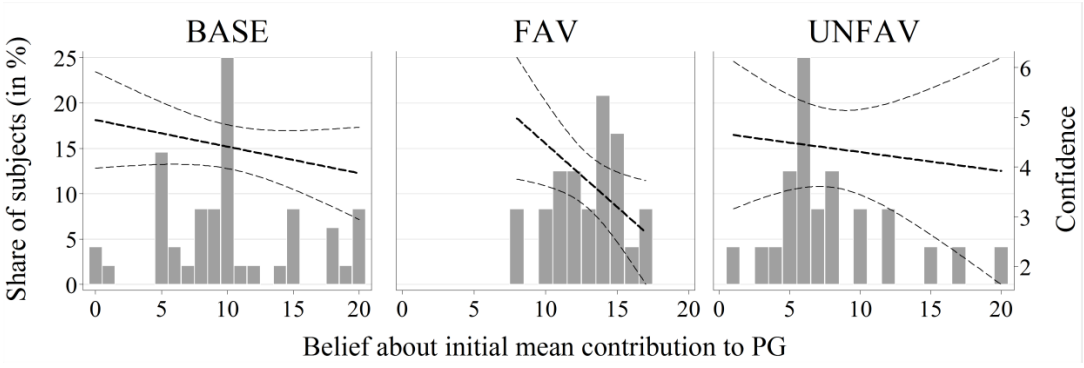


Fig. 6 –Beliefs and Confidence

Notes:Base has 48 subjects, FAV and UNFAV have 24 subjects respectively. The bars denote the share of subjects in a given treatment expressing a certain belief. The dashed lines represent linear prediction plots for the subjects' self-reported Confidence (with 95% confidence intervals). Confidence is measured on a 7-point Likert scale ranging from 1 ("I am very unsure") to 7 ("I am very sure").

As Fig. 6 shows, the information manipulation has a strong effect on beliefs. In Base, subjects expect the first period's contribution to be 10.27 on average. The belief drops to 8.04 in UNFAV, i.e. if they know that participants received the unfavorable information sheet. The difference is significant (ranksum test, p=0.0472, two-sided). By contrast, the average belief increases from 10.27 to 12.96 in FAV. Again, the difference is significant (ranksum test, p=0.0054, two-sided).

Mean confidence does not vary significantly across treatments. Interestingly however, as can be seen in Fig. 6, in each treatment confidence is negatively correlated with beliefs. Subjects with a relatively pessimistic view of other's cooperativeness are considerably more confident than those with optimistic beliefs. This pattern is significant and very robust[9] and might give us an indication for why actual contributions react strongly to negative selective information, but only cautiously to positive selective information. The less confident optimistic beliefs might be more prone to external manipulation than the more confident pessimistic beliefs.

[9]Log-Level OLS regression of ln(confidence) on belief. A 1-point increase in beliefs corresponds on average to a 2 % decrease in confidence (p=0.004). This result is robust to adding treatment dummies, changing the functional form to Level-Level or Log-Log, or running a Tobit.

Taken together, the additional experiments provide further evidence for the above interpretation of our results from the repeated public-good game. If people face a cooperation problem and observe "broken windows", they strongly expect that others are not likely to cooperate, so they themselves also do not cooperate. By contrast, if "windows are fixed", they believe that others might cooperate, but at the same time acknowledge the cooperation dilemma. Thus, they are not very confident in their expectation. This translates into only intermediate rates of cooperation – though this seemingly suffices to at least sustain cooperation over time.

V. Discussion

In this paper, we have explored whether cooperative behavior reacts to selective pre-play information about other, unrelated groups. To this end, we have used an experimental framework that captures cooperation situations which are sufficiently rich, so that several potentially conflicting behavioral norms might emerge. We find that the aggregate level of cooperativeness is dramatically reduced by giving subjects examples of uncooperative behavior. The opposite intervention was less powerful, but still effective. If participants in an otherwise uncooperative environment saw examples of successful groups, cooperation rates were stabilized, while they quickly decayed otherwise. A likely reason for these findings is that cooperation is path-dependent, i.e., long-run behavior is strongly correlated with initial contribution rates – which, in turn, are a reaction to the pre-play information given to the subjects. The latter might be due to a change in subjects' expectations about others' behavior, as was suggested by our additional experiments.

Our findings have a number of important implications. First, they clearly point to the relevance of pre-game communication – a factor which has only lately started to receive significant attention in the literature. While the existing literature usually focuses on self-chosen cheap talk messages (for an overview, see Crawford 1998), we demonstrate that also exogenously selected, one-way information about other players can alter how players act in subsequent games. In particular the findings of our second series of experiments might be of interest to this literature. They suggest a possible channel through which the observed effects of cheap talk are mediated, namely through the alteration of subjects' pre-game expectations.

Second, and closely related, is our finding that there seems to be a strong asymmetry in the strength of the reaction to the mechanism. By giving subjects examples of

uncooperative behavior, the aggregate level of cooperativeness is immediately reduced. By contrast, optimistic foreign experiences need time to unfold their beneficial effect. Bad impressions carry more weight than good ones, so that creating bad behavior is seemingly easier than producing good one – a finding which is also mirrored in previous work on the prevalence of "bad" vs. "good" (cf. Baumeister, Bratslavsky et al. 2001). Thus, one might expect that also other mechanisms which build on the manipulation of beliefs perform differently, depending on the content of the experience as well as on the specific environment in which they occur.

Taken together, our results underline the power and importance of information and experience in shaping cooperative behavior. The bottom line is that observation matters. Interestingly, people do not only learn from what they experience themselves. They also seem to learn "vicariously", by observing others, or by seeing the results (Bandura 1977). The effect is even present when participants are told that the information they are receiving is selective. This connects our study with the growing literature on social learning (for a recent meta-study see Weizsäcker 2010). Our subjects' behavior critically depends on pre-game expectations, which we show to be easy to deteriorate in a complex setting – simply by providing subjects the opportunity of vicarious learning.

This suggests that, in appropriate circumstances, impression management might indeed be a feasible tool to avert, or at least to mitigate, the danger of social dilemmas. Of course this is a paternalistic intervention. But note that conditional cooperators need not even be deceived for the intervention to be successful. All that is required is that they expect the manipulation to matter for a sufficient fraction of the remaining members of their group, be that because they are deceived, or because they are skeptical themselves, but willing to give cooperation a try since the intervention gives all of them one and the same informational starting point.

For policy makers these findings represent both a chance and a peril. If they do not manage to repair broken windows quickly, both literally and metaphorically speaking, chances are a vicious cycle starts. By contrast, if they can induce some to lead others by their socially beneficial example, this strategy may well work. In particular, policy makers would want to prevent (perhaps wrong) pessimistic beliefs from spreading. In any case, we show that home-grown expectations must not be disregarded in order to attain socially desirable outcomes. Consequently in situations where the success of a law depends on the willingness of individual citizens to

cooperate – for instance in areas like waste separation and sustainable water use – government might want to consider a PR campaign in order to create a general atmosphere of cooperativeness within the population.

Managing first impressions might certainly be less effective if the large majority of addressees know better. The intervention requires a sufficient degree of uncertainty. Yet in political reality, quite a few public goods are characterized by deep conceptual and factual uncertainty. Problems like climate change are heavily contested among scientists and not well understood by many. Addressees have to trust expertise. If in the eyes of addressees the underlying social problem is opaque, they are also likely to be uncertain how others will react to it. Most importantly, addressees face behavioral uncertainty whenever they newly enter a community. They do not know local mores, nor do they know how determined the group is when it comes to enforcing them.

The results in this paper should, of course, not be taken as arguments against the importance of elaborated, incentives-altering mechanisms in general. For instance, the effectiveness of our very simple mechanism seems to be asymmetric, since it has a hard time to increase cooperation in an otherwise uncooperative environment. Still, our findings suggest that minimal interventions *can* have a strong behavioral effect. Future research could try to explore the interaction between such simple and other, more complex and intrusive mechanisms. Besides, it might also be informative to study in more detail if certain player "types" (like selfish actors, conditional cooperators, or altruists) are affected differently by mechanisms in the spirit of the one presented here. It is even conceivable that such interventions change the preferences of certain players. Another important, though challenging, next step would certainly be to move into the field and to investigate the effectiveness of means to foster cooperation in a more natural setup. Extending on our work allows for studying these and other interesting aspects in the future.

References

ANDREONI, JAMES (1988). "Why Free Ride? Strategies and Learning in Public Goods Experiments." Journal of Public Economics **37**: 291-304.

ANDREONI, JAMES and RAGAN PETRIE (2004). "Public Goods Experiments without Confidentiality. A Glimpse into Fund-raising." Journal of Public Economics **88**: 1605-1623.

BANDURA, ALBERT (1977). Social Learning Theory. Englewood Cliffs, Prenctice Hall.

BAUMEISTER, ROY, ELLEN BRATSLAVSKY, CATRIN FINKENAUER and KATHLEEN D. VOHS (2001). "Bad is Stronger than Good." Review of General Psychology **5**: 323-370.

BECKENKAMP, MARTIN, CHRISTOPH ENGEL, ANDREAS GLÖCKNER, BERND IRLENBUSCH, HEIKE HENNIG-SCHMIDT, SEBASTIAN KUBE, MICHAEL KURSCHILGEN, ALEXANDER MORELL, ANDREAS NICKLISCH, HANS-THEO NORMANN and EMANUEL TOWFIGH (2009). Beware of Broken Windows! First Impressions in Public-good Experiments, mimeo.

BLUME, ANDREAS and ANDREAS ORTMANN (2007). "The Effects of Costless Pre-Play Communication. Experimental Evidence from Games with Pareto-Ranked Equilibria." Journal of Economic Theory **132**: 274-290.

BRAGA, ANTHONY A., DAVID L. WEISBURD, ELIN J. WARING, LORRAINE GREEN MAZEROLLE, WILLIAM SPELMAN and FRANCIS GAJEWSKI (1999). "Problem-Oriented Policing in Violent Crime Places. A Randomized Controlled Experiment." Criminology **37**: 541-580.

BREKKE, KJELL ARNE, KAREN EVELYN HAUGE, JO THORI LIND and KARINE NYBORG (2009). Playing with the Good Guys: A Public Good Game with Endogenous Group Formation, mimeo.

BURLANDO, ROBERTO M. and FRANCESCO GUALA (2005). "Heterogeneous Agents in Publilc Goods Experiments." Experimental Economics **8**: 35-54.

CAMERON, ADRIAN COLIN and P. K. TRIVEDI (2005). Microeconometrics. Methods and Applications. Cambridge ; New York, Cambridge University Press.

CIALDINI, ROBERT B., RAYMOND R. RENO and CARL A. KALLGREN (1990). "A Focus Theory of Normative Conduct. Recycling the Concept of Norms to Reduce Littering in Public Places." Journal of Personality and Social Psychology **58**: 1015-1026.

CINYABUGUMA, MATTHIAS, TALBOT PAGE and LOUIS PUTTERMAN (2005). "Cooperation under the Threat of Expulsion in a Public Goods Experiment." Journal of Public Economics **89**: 1421-1435.

CORAZZINI, LUCCA, SEBASTIAN KUBE, MICHEL ANDRÉ MARÉCHAL and ANTONIO NICOLO (2010). Elections and Deceptions. Theory and Experimental Evidence

CRAWFORD, VINCENT (1998). "A Survey of Experiments on Communication via Cheap Talk." Journal of Economic Theory **78**(2): 286-298.

CROSON, RACHEL T.A., ENRIQUE FATAS and TIBOR NEUGEBAUER (2008). The Effect of Excludability on Team Production, mimeo.

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

FALK, ARMIN and URS FISCHBACHER (2006). "A Theory of Reciprocity." Games and Economic Behavior **54**: 293-315.

FALKINGER, JOSEF, ERNST FEHR, SIMON GÄCHTER and RUDOLF WINTER-EBMER (2000). "A Simple Mechanism for the Efficient Provision of Public Goods. Experimental Evidence." American Economic Review **90**: 247-264.

FEHR, ERNST and SIMON GÄCHTER (2000). "Cooperation and Punishment in Public Goods Experiments." American Economic Review **90**: 980-994.

FEHR, ERNST and SIMON GÄCHTER (2002). "Altruistic Punishment in Humans." Nature **415**: 137-140.

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 and ERNST FEHR (2001). "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." Economics Letters **71**: 397-404.

FORSYTHE, ROBERT, ROGER B. MYERSON, THOMAS RIETZ and ROBERT J. WEBER (1993). "An Experiment on Coordination in Multi-Candidate Elections. The Importance of Polls and Election Histories." Social Choice & Welfare **10**: 223-247.

FORSYTHE, ROBERT, THOMAS RIETZ, ROGER B. MYERSON and ROBERT J. WEBER (1996). "An Experimental Study of Voting Rules and Polls in Three-Candidate Elections." International Journal of Game Theory **25**: 355-383.

GÄCHTER, SIMON and CHRISTIAN THÖNI (2007). "Social Learning and Voluntary Cooperation Among Like-Minded People." Journal of the European Economic Association **3**: 303-314.

GALBIATI, ROBERTO and PIETRO VERTOVA (2008). "Law and Behaviours in Social Dilemmas. Testing the Effect of Obligations on Commitment." Games and Economic Behavior **64**: 146-170.

GLÖCKNER, ANDREAS, BERND IRLENBUSCH, ANDREAS NICKLISCH and HANS-THEO NORMANN (2010). "Leading With(out) Sacrifice? A Public-Goods Experiment with a Privileged Player." Economic Inquiry ***: ***.

GOERG, SEBASTIAN J. and GARY WALKOWITZ (2010). "On the Prevalence of Framing Effects Across Subject-Pools in a Two-Person Cooperation Game." Journal of Economic Psychology ***: ***.

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

GUNNTHORSDOTIR, ANNA, DANIEL HOUSER and KEVIN MCCABE (2007). "Disposition, History and Contributions in Public Goods Experiments." Journal of Economic Behavior & Organization **62**: 304-315.

GÜRERK, ÖZGÜR, BERND IRLENBUSCH and BETTINA ROCKENBACH (2006). "The Competitive Advantage of Sanctioning Institutions." Science **312**: 108-111.

HERRMANN, BENEDIKT, CHRISTIAN THÖNI and SIMON GÄCHTER (2008). "Antisocial Punishment Across Societies." Science **319**: 1362-1367.

ISAAC, R. MARK, JAMES M. WALKER and ARLINGTON W. WILLIAMS (1994). "Group Size and the Voluntary Provision of Public Goods. Experimental Evidence Utilizing Large Groups." Journal of Public Economics **54**: 1-36.

KEIZER, KEES, SIEGWART LINDENBERG and LINDA STEG (2008). The Spreading of Disorder. Science. **322:** 1681-1685.

KLOR, ESTEBAN and EYAL WINTER (2007). "The Welfare Effects of Public Opinion Polls." International Journal of Game Theory **35**: 379-394.

KOGAN, SHIMON, ANTHONY M. KWASNICA and ROBERTO A. WEBER (2010). "Coordination in the Presence of Asset Markets." American Economic Review ***: ***.

KUBE, SEBASTIAN and CHRISTIAN TRAXLER (2010). "The Interaction of Legal and Social Norm Enforcement." Journal of Public Economic Theory ***: ***.

MAMMEN, ENNO (1993). "Bootstrap and Wild Bootstrap for High Dimensional Linear Models." Annals of Statistics **21**: 255-285.

MENGEL, FRIEDRIKE (2007). "The Evolution of Function-Value Traits for Conditional Cooperation." Journal of Theoretical Biology **245**: 564-575.

NIKIFORAKIS, NIKOS S. (2008). "Punishment and Counter-Punishment in Public Good Games: Can We Really Govern Ourselves?" Journal of Public Economics **92**: 91-112.

OSTROM, ELINOR, JAMES M. WALKER and ROY GARDNER (1992). "Covenants with and without Sword. Self-Governance is Possible." American Political Science Review **40**: 309-317.

PAGE, TALBOT, LOUIS PUTTERMAN and BULENT UNEL (2005). "Voluntary Association in Public Goods Experiments. Reciprocity, Mimicry and Efficiency." Economic Journal **115**: 1032-1053.

RAMOS, JOAO and BENNO TORGLER (2010). Are Academics Messy? Testing the Broken Windows Theory with a Field Experiment in the Work Environment http://www.bepress.com/cgi/viewcontent.cgi?article=1505&context=feem.

WEBER, ROBERTO A. (2006). "Managing Growth to Achieve Efficient Coordination in Large Groups." American Economic Review **96**: 114-126.

WEIZSÄCKER, GEORG (2010). "Do We Follow Others When We Should? A Simple Test of Rational Expectations." American Economic Review ***.

WILSON, JAMES Q. and GEORGE L. KELLING (1982). "Police and Neighborhood Safety. Broken Windows." Atlantic Monthly **127**: 29-38.

WU, C.F.J. (1986). "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis." Annals of Statistics **14**: 1261-1295.

ZIMBARDO, PHILIP (1969). "The Human Choice. Individuation, Reason, and Order versus Deindividuation, Impulse, and Chaos." Nebraska Symposium on Motivation **17**: 237-307.

ZIMBARDO, PHILIP G. and EBBE B. EBBESEN (1969). Influencing Attitudes and Changing Behavior. A Basic Introduction to Relevant Methodology, Theory, and Applications. Reading, Mass.,, Addison-Wesley Pub. Co.

Appendix 1: Instructions Baseline

You are now taking part in an economic experiment. If you read the following instructions carefully, you can, depending on your decisions, earn a considerable amount of money. It is therefore important that you take your time to understand the instructions.

The instructions which we have distributed to you are for your private information. **Please do not communicate with the other participants during the experiment.** Should you have any questions please ask us.

During the experiment we shall not speak of Pounds, but of Experimental Currency Units (ECU). Your entire earnings will be calculated in ECUs. At the end of the experiment the total amount of ECUs you have earned will be converted to Pounds at the rate of **1 ECU = 4 p** and will be immediately paid to you in cash.

At the beginning of the experiment the participants will be randomly divided into groups of four. You will therefore be in a group with 3 other participants**. The composition of each group will remain the same throughout the experiment.** The experiment lasts 10 periods and each period is divided into 3 stages.

The 1st stage:

At the beginning of each of the 10 periods each participant will receive 20 ECUs. In the following, we shall refer to this amount as the "endowment". In the 1st stage, your task is to decide how to use your endowment. You have to decide how many of the 20 ECUs you want to contribute to a project (from 0 to 20) and how many of them to keep for yourself. The consequences of your decision are explained in detail below.

Once all the players have decided their contribution to the project you will be informed about, the group's total contribution, your income from the project and your payoff in this period. Your payoff in each period is calculated using the following simple formula. Again, if you have any difficulties do not hesitate to ask us.

Income from the 1st stage	=	Endowment of ECUs	–	Your contribution to the Project	+	0.4*Total contribution to the Project

This formula shows that your 1st stage income consists of two parts:

1) The ECUs which you have kept for yourself (endowment – contribution)
2) The income from the project, which equals to the 40% of the group's total contribution.

The income of each group member from the project is calculated in the same way. This means that each group member receives the same income from the project. Suppose the sum of the contributions

of all group members are 60 ECUs. In this case, each member of the group receives an income from the project of: 0.4*60=24 ECUs. If the total contribution to the project is 9 points, then each member of the group receives an income of: 0.4*9=3.6 ECUs from the project.

You always have the option of keeping the ECUs for yourself or contributing them to the project. Each ECU that you keep raises your end of period income by 1 ECU. Supposing you contributed this point to the project instead, then the total contribution to the project would rise by 1 ECUs. Your income from the project would thus rise by 0.4*1=0.4 ECUs. However, the income of the other group members would also rise by 0.4 ECUs each, so that the total income of the group from the project would be 1.6 points. Your contribution to the project therefore also raises the income of the other group members. On the other hand you also earn an income for each point contributed by the other members to the project. In particular, for each point contributed by any member you earn 0.4 ECUs.

In addition to the 20 ECUs per period, each participant receives a one-off lump sum payment of 25 ECUs at the beginning of this part. This one-off payment can be used to pay for eventual losses during the experiment. **However, you can always evade losses with certainty through your own decisions.** Note that this lump sum payment will not be used to calculate the income from the period. It will only be added to your total income from all the periods at the very end.

The 2nd stage:

At the 2nd stage you will be informed how much each group member contributed individually to the project at the 1st stage. At this stage you can **reduce or leave equal** the income of **each** member of your group by **distributing points**. The other group members can also reduce your income if they wish to.

If you choose 0 points for a particular group member, you do not change his or her income. However if you give a member 1 point you reduce his or her income by 10 percent. If you give a member 2 points you reduce his or her income by 20 percent, etc. The amount of points you distribute to each member determines, therefore, how much you reduce their income from the 1st stage. If one player receives in total 4 points his income will be reduced by 40% and if he receives **10 or more** his income from the 1st stage will be reduced by 100%.

If you distribute points you have costs in ECUs, which depend on the amount of points you distribute. You can distribute between 0 and 10 points to each group member. The more points you give **to any** group member, the higher your costs. Your total costs are equal to the **sum of the costs of distributing points to each of the other three group members**. The following table illustrates the relation between distributed points to **each** group member and the cost of doing so in ECUs.

Points	0	1	2	3	4	5	6	7	8	9	10
Cost of points per person	**0**	**1**	**2**	**4**	**6**	**9**	**12**	**16**	**20**	**25**	**30**

Example: Supposing you give 2 points to player 1 this costs you 2 ECUs; if you also give 8 points to player 3 this costs you a further 20 ECUs; and if you give 0 points to the last group member this has no additional cost for you. In this case, your total costs of distributing points would be 22 ECUs (2+20+0) and not 30 ECUs.

The following equation summarizes the previous information. Your total income from the two stages is calculated as follows:

Income at the end= of the 2nd stage	Income from the* 1st stage	[(10 – received– points)/10]	Costs of distributed points

Please note that your income in ECUs at the end of the 2nd and the 3rd stage can be negative, if the costs of your points distributed exceeds your (possibly reduced) income from the 1st stage. **You can however evade such losses with certainty through your own decisions.** Should your income become zero or negative at the end of the 2nd stage you will not be able to continue to the 3rdstage. If your income becomes zero or negative at the end of the 3rd stage you can simply use your 25 ECUs that we gave you in the beginning in order to pay this off.

The 3rd stage:

In the 3rd and final stage, after being informed of the points that the other group members assigned to you, you will be given one last opportunity of assigning points back to the other participants, thus reducing their income. We shall call these points "counter-points". **You will only be able to assign counter-points to participants who assigned points to you during the 2nd stage**

The costs of assigning points, as well as the income reduction caused by each point remain the same as before. The following table shows you how to calculate the costs for assigning counter-points.

Costs for assigning **counter-points** to <u>one specific group member</u> in the **3rd stage**											
Points that you already assigned to <u>one specific group member</u> in the **2nd stage**	**Counter-points** you assign to that <u>same group member</u> in the **3rd stage**										
	0	1	2	3	4	5	6	7	8	9	10
0	**0**	**1**	**2**	**4**	**6**	**9**	**12**	**16**	**20**	**25**	**30**
1	**0**	**1**	**3**	**5**	**8**	**11**	**15**	**19**	**24**	**29**	
2	**0**	**2**	**4**	**7**	**10**	**14**	**18**	**23**	**28**		
3	**0**	**2**	**5**	**8**	**12**	**16**	**21**	**26**			
4	**0**	**3**	**6**	**10**	**14**	**19**	**24**				
5	**0**	**3**	**7**	**11**	**16**	**21**					
6	**0**	**4**	**8**	**13**	**18**						
7	**0**	**4**	**9**	**14**							
8	**0**	**5**	**10**								
9	**0**	**5**									
10	**0**										

Example: if you distribute 2 points in the 2nd stage to player 1 you have a cost of 2 ECUs If in the 3rd stage you decide to distribute 3 counter-points to player 1, a further 7 ECUs are added to your cost.

Your income after the 3rd stage (= period income) is therefore calculated as follows:

Income at the end= of the 3rd stage	Income from the* 2nd stage	[(10 – received counter-points)/10]	– Cost of distributed counter-points

If you have any further questions please raise your hand and one of the supervisors

Appendix 2: Control Questionnaire

1. Each group member has an endowment of 20 ECUs. Nobody (including you) contributes any ECUs to the project. What is:
   a. Your income at the end of the first stage? ……
   b. The income of the other group members?……

2. Each group member has an endowment of 20 ECUs. You contribute 20 ECUs to the project. All other group members contribute 20 ECUs each to the project. What is:
   a. Your income at the end of the first stage? ……
   b. The income of the other group members?……

3. Each group member has an endowment of 20 ECUs. The other three group members contribute together a total of 30 ECUs to the project. What is:
   a. Your income at the end of the first stage if you contribute 0 ECUs to the project? …….
   b. Your income at the end of the first stage if you contribute 15 ECUs to the project? …….

4. Each group member has an endowment of 20 ECUs. You contribute 8 ECUs to the project. What is:
   a. Your income at the end of the first stage if the other group members together contribute a further total of 7 ECUs to the project?.....
   b. Your income at the end of the first stage if the other group members together contribute a further total of 22 ECUs to the project?.....

5. At the second stage you distribute the following points to your three other group members: 9, 5, 0. What are the total costs of your distributed points?….

6. What are your costs if you distribute 0 points? ……

7. By how many percent will your income from the first stage be reduced when you receive from the other group members a total of:
   a. 0 points? …
   b. 4 points? ...
   c. 15 points? …

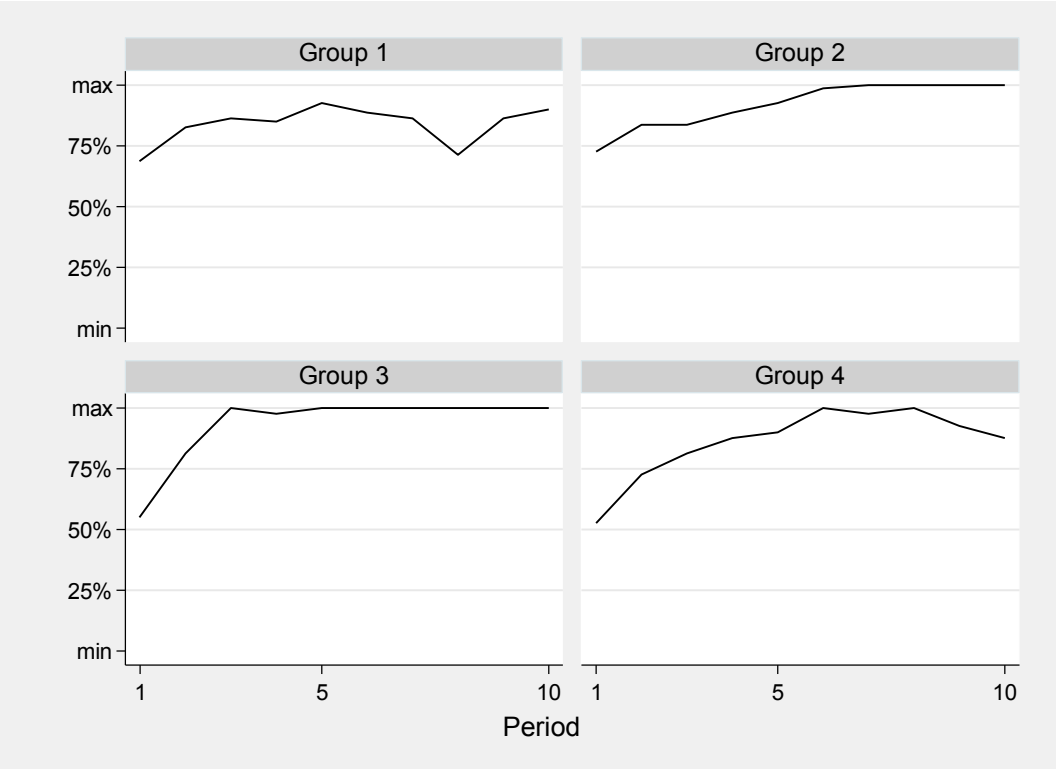
8. At the second stage you distribute the following points to your three other group members: 2, 2, 0. In the third stage you distribute the following points to your three other group members: 1, 1, 1. What are the total costs of your distributed points?….

9. By how many per cent is your second stage income reduced, if you have received the sum of 3 counter-points from the other group members in stage 3?

Appendix 3: Additional Sheet FAV Treatment

Additional information for today's experiment

This experiment has already been run at two laboratories in Bonn (Germany) and London (UK). The following figure shows you the contribution behaviour in four selected groups from both locations.



In these selected graphs you see how much the 4 group members contributed on average to the group project. In the selected groups 1 and 2 the contributions are high right from the beginning. In the selected groups 3 and 4, average contribution starts somewhat lower but rises over the course of the experiment.

Some additional numbers from the experiments in Bonn and London:

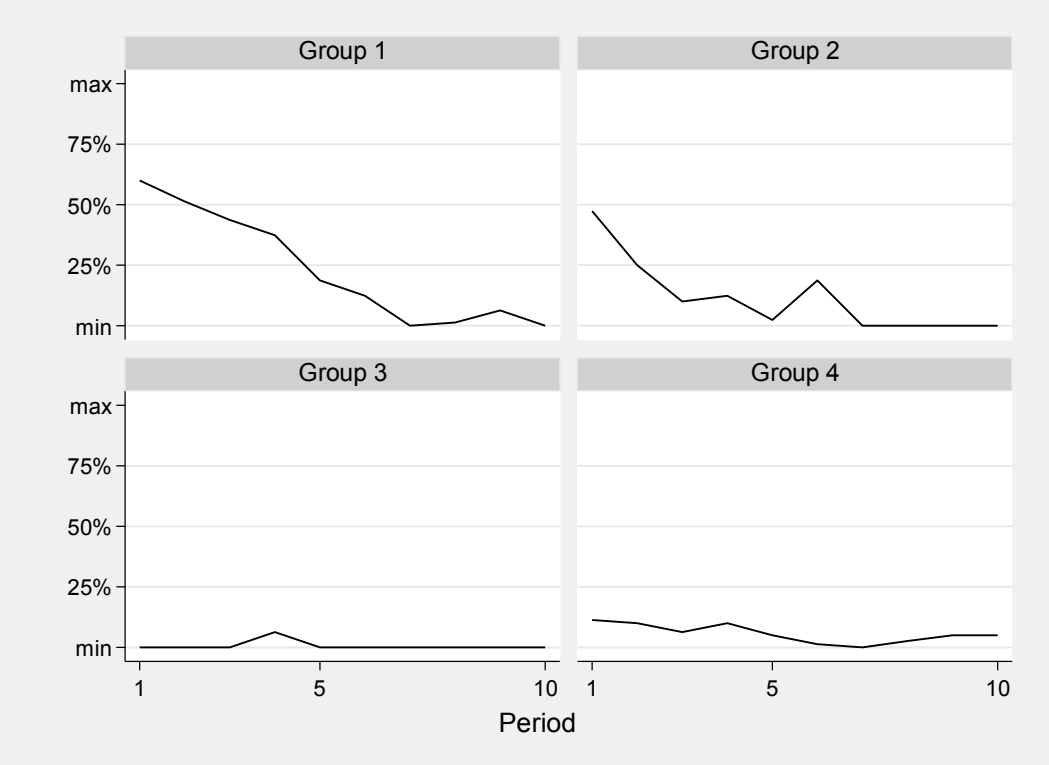
a) In the first round of the two previous experiments, 58% to 74% of the contributions were between 10 and 20.
b) In the last round, up to 53% of the contributions were equal to 20.
c) Players, whose income was reduced in stage 2, only used counter-points in stage 3 in 29% to 41% of all possible cases.

All these graphs and numbers are meant to give you some orientation concerning the type of situation you might be in during the experiment and what you might possibly expect from the other members of your group.

Appendix 4: Additional Sheet UNFAV Treatment

Additional information for today's experiment

This experiment has already been run at two laboratories in Bonn (Germany) and London (UK). The following figure shows you the contribution behaviour in four selected groups from both locations.



In these selected graphs you see how much the 4 group members contributed on average to the group project. In the selected groups 1 and 2 the contributions are rather low and display a decrease over the course of the experiment. In the selected groups 3 and 4, average contributions are low right from the beginning

Some additional numbers from the experiments in Bonn and London:

a) In the first round of the two previous experiments, 37% to 55% of the contributions were between 0 and 10.
b) In the last round, up to 52% of the contributions were equal to 0.
c) Players, whose income was reduced in stage 2, only used counter-points in stage 3 to reduce the income of those who had given them points in stage 2. Every time they used counter-points they reduced the income of those who had given them points in stage 2 by 18% to 25%.

All these graphs and numbers are meant to give you some orientation concerning the type of situation you might be in during the experiment and what you might possibly expect from the other members of your group.

Appendix 5: Belief Elicitation

As noted in the main text, subjects in the belief elicitation had to read the same instructions and answer the same control questions as the participants in the contribution treatments. However, once ztree started subjects were told the following on their screens:

You have just read the instructions and answered the control questions. This game has already been played in other laboratories. The participants of those experiments read the same instructions as you and answered the same control questions. You are not going to play this game today. Instead, your task is to answer the following question:

What do you think, how much did those previous participants contribute on average in the first round to the project?

Your guess……

We have rounded the average down to the next whole number, i.e. the correct answer is a whole number between 0 and 20. If you guess is exactly correct, you will receive 4 Euros. If your guess is wrong, you will receive 0 Euros.

Appendix 6: Additional Data Analysis

Table A1: Summary Statistics of Punishment and Counter-punishment across Treatments

	BASEFAV	BASEUNFAV	FAV	UNFAV
(1) Punishment	0.65	0.35	0.92	0.51
(2) Counter-punishment	0.17	0.11	0.21	0.23
(3) Counter-punishment (conditional)	0.72	0.78	0.59	1.06

Notes: All figures are means of (counter-)punishment points given per period and subject. (3) is conditional on subjects being able to allocate counter-punishment. Subjects could only allocate counter-punishment in stage 3 if they had been punished in stage 2.

Parametric Analysis

We have participants interacting over 10 announced periods in fixed groups of four. Moreover all our data is right-censored since many participants contribute their entire endowment of 20. Most of our data is also left-censored in that a substantial fraction has kept the entire endowment. The ideal functional form would therefore be a mixed Tobit model. Unfortunately, such a model is not available. Random effects Tobit would ignore that observations within groups are not independent. A linear mixed model would ignore that observations are censored. We therefore estimate Tobit models, with lower level 0 and upper level 20, and correct standard errors via bootstrapping, with resampling at the level of entire groups. All models work with 50 repetitions, which is known to be sufficient for correcting standard errors.

In the comparison between BASEUNFAV and UNFAV, a non-parametric bootstrap (resampling pairs of dependent and independent variables) yields significant effects. We therefore do not report results from a (semi-)parametric bootstrap resampling residuals (for background see Cameron and Trivedi 2005:360 f.), nor from „wild“

bootstrapping, which randomly assigns error terms using the procedure developed by (Wu 1986; Mammen 1993), using the parameters proposed by (Cameron and Trivedi 2005:377); unsurprisingly, showing significance becomes even easier if one relies on these methods. By contrast, these more advanced methods make a difference in the comparison between BASEFAV and FAV, which is why we report all three estimates. Since we only use bootstrapping for correcting standard errors for the dependence of observations, parametric bootstrapping is justified.

In the comparison of BASEUNFAV with UNFAV, in all models we find a strong and highly significant negative effect of the belief manipulation. There is no time trend, and no interaction between time and treatment. By contrast, punishment has a highly significant negative main effect: those who have been punished in the previous period still contribute less than the group average in the subsequent period. This holds irrespective of the belief manipulation, as shown by the insignificance of the interaction effect.

Table A2: Comparing BASEUNFAV with UNFAV
Dependent variable: Contribution

	(1)	(2)	(3)	(4)	(5)	(6)
unfav	-10.708	-10.709	-9.275	-10.972	-11.583	-13.768
	(.001)	(.001)	(.001)	(.001)	(.004)	(.005)
period		.027	.163	-.421	-.474	-.615
		(.896)	(.568)	(.116)	(.215)	(.184)
unfav*period			-.263		.102	.304
			(.453)		(.796)	(.498)
lagged received punishment (lrecpun)				-1.947	-1.949	-3.400
				(.013)	(.013)	(.006)
unfav*lrecpun						2.273
						(.172)
cons	22.104	21.957	21.218	26.022	26.347	27.723
	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)

Note: Tobit, with lower level 0 and upper level 20. Models correct for the dependence of observations per individual and within groups by bootstrapping, with random draws of entire groups. A non-parametric bootstrap is administered, i.e. pairs of dependent and their independent variables are randomly selected. p-values in parenthesis. N = 1152. 95 left censored, 494 right censored observations.

In the comparison of BASEFAV with FAV we get a different picture. We have a positive main effect of the belief manipulation in all models. But it is significant only with wild bootstrapping (models 1, 2, 4), and weakly significant with model 4 (which controls for lagged received punishment). There is, however, a significant interaction between the treatment and the time trend in model 3. It also is present in wild bootstrapping in models 5 and 6 at conventional significance levels, and it is weakly significant with alternative bootstrapping methodologies. This indicates that the belief manipulation stops contributions from quickly eroding. In models 4 and 5, we also find the negative effect of lagged received punishment, as in the comparison of BASEUNFAV with UNFAV.

Table A3: Comparing Basefav with fav
Dependent variable: Contribution

	(1)	(2)	(3)	(4)	(5)	(6)
fav	6.151	6.126	1.071	7.275	3.114	2.612
	(.122)	(.123)	(.789)	(.085)	(.502)	(.662)
	(.211)	(.181)	(.745)	(.074)	(.332)	(.433)
	(.004)	(.006)	(.683)	(.001)	(.286)	(.446)
period		-.584	-1.114	-.939	-1.335	-1.348
		(.010)	(.001)	(<.001)	(<.001)	(.001)
		(.001)	(<.001)	(<.001)	(<.001)	(<.001)
		(<.001)	(<.001)	(<.001)	(<.001)	(<.001)
fav*period			.925		.695	.726
			(.015)		(.092)	(.101)
			(.019)		(.081)	(.072)
			(<.001)		(.014)	(.018)
lagged received punishment				-1.395	-1.369	-1.610
				(.038)	(.046)	(.222)
				(.024)	(.021)	(.101)
				(<.001)	(<.001)	(.003)
fav*lrecpun						.405
						(.791)
						(.709)
						(.560)
cons	7.714	10.906	13.790	13.867	16.205	16.444
	(.023)	(.002)	(<.001)	(.002)	(.001)	(.002)
	(.002)	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)
	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)

Note: Tobit with lower level 0 and upper level 20. Models correct for the dependence of observations per individual and within groups by bootstrapping, with random draws of entire groups. We compare three different bootstrapping techniques, with p-values in parenthesis and in the following order: non-parametric; parametric over residuals; wild bootstrap. N = 1008. 214 left censored, 277 right censored observations.

Chapter 4

Follow the crowd or my conscience?
An experimental study of information, self-reflection, and social preferences

Michael Kurschilgen

Abstract

In a modified dictator game, I test experimentally the robustness of social preferences to (1) information about others' behavior and (2) self-reflection about the morally right behavior. I find that whereas information has virtually no effect, self-reflection changes social preferences substantially. Information does have, however, a strong effect on people's moral judgments. Interestingly, since people differ with respect to their normative goals (social welfare vs. equality), self-reflection leads to more heterogeneity of preferences as subjects' home-grown normative concerns are reinforced. The results can be well explained within the framework of identity utility (Akerlof & Kranton 2000).

Keywords: Social Norms, Morality, Conformity, Identity, Social Preferences, Information, Dictator Game, Experiment

JEL: *C91, D83, D84*

I. Introduction

I love competing. Whenever I play basketball with my brother, it is a really hard fight. None of us wants to lose. So we play tough defense, hustle for every ball and do not shy away from fouling each other. How does it come then that when I play basketball with my little nephew, I do not defend, do not hustle and do not foul? Well, I am not supposed to do so. I am supposed to let him win. In their recent book, Akerlof & Kranton (2010) reason that the crucial difference between two such situations are norms. Playing against my brother, I am subject to a different set of norms than playing against my little nephew. As a consequence my identity changes; I do not perceive myself as a competitor but rather as a playfellow. This in turn changes my preferences. While I very much enjoy winning against my brother, I would actually feel bad if I won against my little nephew.

The concept of identity utility proposed by Akerlof & Kranton (2000 and 2010) captures exactly such preference changes due to contextual shifts of the relevant norms. This contrasts sharply with the predominant notion in modern economics that preferences are inherent characteristics of the individual. Notably, the literature on social preferences (Fehr & Schmidt 1999, Bolton & Ockenfels 2000, Andreoni & Miller 2002, Charness & Rabin 2002) acknowledges that different people might have different preferences for distributive justice but it does not allow the same people to vary their preferences from one situation to another as a result of a change in the relevant norms.

In most everyday situations we know which norm applies and adapt our behavior automatically and unconsciously. But often, we face new situations and do not know what the *right* behavior is. We find guidance in norms. There are two distinct sources from which to deduce a norm. Externally, by watching how others behave ("conformity"), and internally, by reflecting what oneself would judge as ideal behavior ("morality"). Conformity requires information about others' behavior. Morality requires self-reflection about one's own moral standards.

While the idea of self-reflection is rather strange to modern economics, it is not new to the discipline. In fact, it can be traced back to Adam Smith's Theory of Moral Sentiments (Smith 1790), who in turn was inspired by the religious concept of conscience. In the context of simple dictator games Bicchieri & Xiao (2009) and Krupka & Weber (2009) show that both information and self-reflection may influence

giving behavior. This paper is the first to study the effect of information and self-reflection on the broader spectrum of social preferences.

The experimental paradigm used is a modified dictator game (MDG). The MDG asks participants for their preferences in a 2x2 world: you are either richer or poorer than the other player; you can either create income for the other player or destroy it. A MDG in a university lab is clearly not an everyday situation. It thus seems plausible to assume that participants have doubts about the normatively appropriate behavior, which is exactly where normative guidance becomes relevant. In the experimental treatments, subjects are guided externally, by giving them information about representative behavior of previous participants ("*Info* treatment"), and internally, by asking them to make a moral judgment prior to their incentivized choice ("*Reflect* treatment").

Previous studies with similar games (see most recently Iriberri & Rey 2011) show that most people who deviate from material selfishness are guided by either social welfare or equality concerns. The results of this experimental study suggest that self-reflection pushes preferences further away from selfishness and towards one of those two norms. When they are richer than the other player, subjects in the *Reflect* treatment create considerably more income for the other player than in the *Baseline*. Strikingly, when deciders are poorer than recipients, self-reflection makes preferences more heterogeneous. While subjects concerned with social welfare create more, equality seeking players destroy more. On the other hand, information has virtually no effect on revealed social preferences. It does however have a strong effect on participants' moral judgments.

The next section introduces the experimental design. Subsequently, I discuss the related theoretical and experimental literature and derive behavioral predictions. Section four presents the experimental results. Section five concludes the paper.

II. Experimental Design

The *Baseline* is a modified dictator game (MDG) similar to the one of Iriberri & Rey (2011). It is meant to elicit social preferences according to the social utility function by Charness & Rabin (2002). Utility according to Charness & Rabin is a weighted average of my own payoff and the payoff of another person. The weight for the other

person's income is called ρ when my payoff is larger than hers and σ when my payoff is smaller:

$$U_{me} = \rho\pi_{other} + (1 - \rho)\pi_{me} \quad \text{if} \quad \pi_{me} \geq \pi_{other} \tag{1}$$

$$U_{me} = \sigma\pi_{other} + (1 - \sigma)\pi_{me} \quad \text{if} \quad \pi_{me} \leq \pi_{other} \tag{2}$$

This utility function distinguishes four behavioral types. Selfish players, who are only interested in their own material payoff, have $\rho = \sigma = 0$. Social welfare maximizers are characterized by both positive ρ and positive σ. Equality maximizers, who like the payoff difference between themselves and other people to be small, are described by a positive ρ and negative σ. Finally, players who strive to increase the payoff difference between themselves and other persons, so-called competitive types, are characterized by negative ρ and negative σ.

More generally, I prefer an allocation B to an allocation A if $U_{me}^{B} \geq U_{me}^{A}$. Assume I am richer than the other person.[1] Inserting (1) and rearranging one obtains:

$$\pi_{me}^{B} - \pi_{me}^{A} \geq \rho((\pi_{me}^{B} - \pi_{me}^{A}) - (\pi_{other}^{B} - \pi_{other}^{A})) \tag{3}$$

which can be rewritten as:

$$\Delta_{me} \geq \rho(\Delta_{me} - \Delta_{other}) \tag{4}$$

Fixing $\Delta_{me} = -10$ the parameter space reduces to the four situations depicted in Table 1.

Table 1: Parameter Space of the MDG

	$\pi_{me} \geq \pi_{other}$	$\pi_{me} \leq \pi_{other}$
$\Delta_{other} > -10$	$\rho \geq \frac{10}{10 + \Delta_{other}} > 0$	$\sigma \geq \frac{10}{10 + \Delta_{other}} > 0$
$\Delta_{other} < -10$	$\rho \leq \frac{10}{10 + \Delta_{other}} < 0$	$\sigma \leq \frac{10}{10 + \Delta_{other}} < 0$

[1] If I am poorer than the other person, ρ is replaced by σ.

The MDG devotes one decision panel to each of these four situations. This is illustrated in Table 2. In the two panels on the left, the decider's payoff is always higher than the recipient's ($\pi_{me} \geq \pi_{other}$) whereas in the two right-hand panels it is the other way around ($\pi_{me} \leq \pi_{other}$). In the two upper panels the decider can create income for the recipient ($\Delta_{other} > 0 > -10$) at a fixed cost of 10 tokens whereas in the two lower panels she can destroy ($\Delta_{other} < -10$) the recipient's income at fixed cost of 10. There are nine decision tasks per panel. In each task the decider has to choose between Option A and Option B, specifying two different payoff allocations for the decider and the corresponding recipient. Option A is the same for every task within a given panel. Option B creates or destroys income of the recipient. Take for example task 1 of the Ahead-Create panel. If the decider chooses Option A she receives 170 tokens and the recipient 70 tokens and if she chooses Option B she gets 160 and the recipient 82.

In every panel the relative price of creating/destroying decreases with every task. In task 1, the decider has to give up 10 tokens to create/destroy 12 tokens whereas in task 9 for the same cost the decider creates/destroys 84 tokens. Consequently, choosing Option B in task 1 and Option A in task 2 of the same panel would violate the General Axiom of Revealed Preferences (Afriat 1967, Varian 1982). In the MDG, a GARP-consistent decider should have at most one switch from Option A to Option B per panel, and no switch from B to A. In addition, consistency requires players not to both create and destroy when they are ahead (or behind). If these consistency requirements are met, the ρ and σ of a given decider are defined by the point at which she switches from Option A to Option B.

For example, a player who chooses Option A in the first 3 tasks of the Ahead-Create panel and Option B in the remaining 6 tasks, would have $0.36 > \rho \geq 0.29$. The same player might then for instance choose always Option A in the Ahead-Destroy panel and in the Behind-Create panel but then switch to Option B in task 7 of the Behind-Destroy panel. This would yield $-0.31 < \sigma \leq -0.23$. The type classification is straightforward: Selfish players will never choose Option B since this is costly. Social Welfare Maximizers will create income for the recipient both when ahead and when behind as long as the relative price of creating is low enough. Equality maximizers will also create when ahead but destroy when behind. Competitive types will destroy recipients' income no matter whether they are ahead or behind.

Table 2: Decision Panels and Tasks of the MDG

		Ahead					Behind				
Create		π^A_{me}=170, π^A_{other}=70					π^A_{me}=110, π^A_{other}=120				
	Task	π^B_{me}	π^B_{other}	Δ_{me}	Δ_{other}	$\rho \geq$	π^B_{me}	π^B_{other}	Δ_{me}	Δ_{other}	$\sigma \geq$
	1	160	82	*-10*	*12*	*0.45*	100	132	*-10*	*12*	*0.45*
	2	160	84	*-10*	*14*	*0.42*	100	134	*-10*	*14*	*0.42*
	3	160	88	*-10*	*18*	*0.36*	100	138	*-10*	*18*	*0.36*
	4	160	94	*-10*	*24*	*0.29*	100	144	*-10*	*24*	*0.29*
	5	160	102	*-10*	*32*	*0.24*	100	152	*-10*	*32*	*0.24*
	6	160	112	*-10*	*42*	*0.19*	100	162	*-10*	*42*	*0.19*
	7	160	124	*-10*	*54*	*0.16*	100	174	*-10*	*54*	*0.16*
	8	160	138	*-10*	*68*	*0.13*	100	188	*-10*	*68*	*0.13*
	9	160	154	*-10*	*84*	*0.11*	100	204	*-10*	*84*	*0.11*
Destroy		π^A_{me}=140, π^A_{other}=130					π^A_{me}=90, π^A_{other}=180				
	Task	π^B_{me}	π^B_{other}	Δ_{me}	Δ_{other}	$\rho \leq$	π^B_{me}	π^B_{other}	Δ_{me}	Δ_{other}	$\sigma \leq$
	1	130	118	*-10*	*-12*	*-5.00*	80	168	*-10*	*-12*	*-5.00*
	2	130	116	*-10*	*-14*	*-2.50*	80	166	*-10*	*-14*	*-2.50*
	3	130	112	*-10*	*-18*	*-1.25*	80	162	*-10*	*-18*	*-1.25*
	4	130	106	*-10*	*-24*	*-0.71*	80	156	*-10*	*-24*	*-0.71*
	5	130	98	*-10*	*-32*	*-0.45*	80	148	*-10*	*-32*	*-0.45*
	6	130	88	*-10*	*-42*	*-0.31*	80	138	*-10*	*-42*	*-0.31*
	7	130	76	*-10*	*-54*	*-0.23*	80	126	*-10*	*-54*	*-0.23*
	8	130	62	*-10*	*-68*	*-0.17*	80	112	*-10*	*-68*	*-0.17*
	9	130	46	*-10*	*-84*	*-0.14*	80	96	*-10*	*-84*	*-0.14*

Note: To ensure that stakes are comparable across panels, every panel has approximately the same average pie size $\bar{P} = \frac{1}{9}\sum_{i=1}^{9}\left(\frac{1}{2}\left(\pi^A_{me,i} + \pi^A_{other,i}\right) + \frac{1}{2}\left(\pi^B_{me,i} + \pi^B_{other,i}\right)\right)$. Ahead-Create has 127 tokens, Ahead-Destroy 123, Behind-Create 122, and Behind-Destroy 123.

In the *Baseline*, after reading the experimental instructions (see Appendix A), one half of the subjects is randomly allotted the role of decider and the other half the role of recipient. On their computer screens, subjects are informed about their role in the experiment. Every decider is then shown on the computer screen, successively, the

four decision panels in which she has to choose her preferred allocation (Option A or Option B) for each of the MDG's 36 tasks, thereby determining their own payoffs and the payoffs of the corresponding recipients.

The *Info* treatment provides subjects with opportunity to consider external normative guidance. After reading the instructions, subjects are informed that the game has been run before with more than 100 deciders. When playing the MDG they see on their screens for each of the 36 tasks which percentage of deciders in the *Baseline* preferred Option A.[2]

The *Reflect* treatment nudges subjects to look for internal normative guidance. After reading the instructions but before allotting the roles of decider and recipient, subjects are asked for their moral judgments. Specifically, they have to state for each of the 36 tasks the will be seeing in the MDG: "Which of the two Options (A or B) do you find morally right?" The instructions on the computer screen make it clear that the answers to this question are not payoff relevant and will not be revealed to other participants. Hereafter, subjects are allotted their roles and play the payoff-relevant MDG. When playing the MDG they see on their screens for each of the 36 tasks their own moral judgments.

The experiment was conducted at the EconLab in Bonn, Germany with a total of 544 participants (272 deciders). There were 304 participants (152 deciders) in the *Baseline*, 144 (72) in the *Info* treatment, and 96 (48) in the *Reflect* treatment. Subjects were recruited randomly per email from a pool of more than 5000 people, using the software ORSEE (Greiner, 2004). The experiment was computerized in ztree (Fischbacher, 2007). Subjects were paid in cash at the end of the experiment after the computer had randomly selected one payoff-relevant task out of each of the four decision panels. The exchange rate was 100 tokens = € 1. The experiment lasted on average 20 minutes and participants earned on average € 6 (ca. US$ 8).

III. Related Literature and Behavioral Predictions

Iriberri & Rey (2010 and 2011) present a similar MDG to elicit social preferences. The two main differences are that, first, Iriberri & Rey have deciders choose between three options (selfish vs. create vs. destroy) while in this experiment they only choose

[2] See Appendix C for the exact percentages in every decision task.

between two (selfish vs. create/destroy) and, second, Iriberri & Rey present the tasks randomly whereas I sort them into four panels and within every panel by the relative price of creating/destroying. The present design aims for subjects' to make choices that are deliberate and conscious and not due to confusion or cognitive overload. As a consequence, it should thus be easier for subjects to choose their answers consistent with the GARP. A problem with the data of Iriberri & Rey is that less than half of the subjects' preferences can be elicited directly from the data since answers are not consistent. As a consequence, Iriberri & Rey have to rely heavily on maximum likelihood estimation, and in many instances subjects' social preferences cannot be clearly determined.

Just as Iriberri & Rey, I measure people's social preferences according to the model of Charness & Rabin (2002). There are some reasons to prefer Charness & Rabin over other models of social preferences discussed in the literature. The model of Charness & Rabin is functionally equivalent to Fehr & Schmidt (1999) but Charness & Rabin place fewer restrictions on the range of plausible parameter values than Fehr & Schmidt.[3] As a result Charness & Rabin can also accommodate social welfare maximizing and competitive types besides the selfish and equality maximizing (i.e. inequality averse) types of Fehr & Schmidt. Bolton & Ockenfels (2000) propose a slightly different conceptualization of equality concerns, which focuses on relative payoff differences instead of absolute differences like Fehr & Schmidt and Charness & Rabin. But just as Fehr & Schmidt their model does not allow for social welfare concerns.

Andreoni & Miller (2002) introduce a more general functional form for social preferences. Their constant elasticity of substitution (CES) function not only allows π_{me} and π_{other} to be perfect substitutes, as both Fehr & Schmidt and Charness & Rabin assume, but also Leontief, Cobb-Douglas and many other.[4] However, the increase in generality comes along with an additional free parameter (i.e. the elasticity of substitution) that needs to be estimated empirically and might vary

[3] One can simply rewrite (1) as $U_{me} = \pi_{me} - \rho(\pi_{me} - \pi_{other})$ and (2) as $U_{me} = \pi_{me} + \sigma(\pi_{other} - \pi_{me})$ and replace $\rho = \beta$ and $\sigma = -\alpha$. Whereas Fehr & Schmidt assume $0 \leq \beta < 1$, Charness & Rabin allow ρ to be negative (competitive types). Similarly, while Fehr & Schmidt assume $0 \leq \alpha$, Charness & Rabin allow ρ to be positive (social welfare maximizers).

[4] Using the same terminology as above, social utility according to Andreoni & Miller can be written as: $U_{me} = (\rho {\pi_{other}}^{s} + (1 - \rho){\pi_{me}}^{s})^{1/s}$. For $s = 1$ the expression simplifies to Charness & Rabin.

substantially across individuals. To fit Andreoni & Miller with a similar MDG we would need numerous additional decision tasks.

An alternative and very promising way to elicit people's social preferences is proposed by Fisman et al. (2007). They use a graphical interface to confront subjects with 50 different allocation problems in which deciders have to choose a point on a two-dimensional graph that represents a budget set. On the one hand the Fisman et al. (2007) approach is clearly superior to both Iriberri & Rey and this paper's. When a subject chooses Option A over B (over C) we can deduce that she prefers A over one (two) other possible allocations. But when a subject chooses a specific point on a two-dimensional budget set we can deduce that she prefers that specific allocation to all other feasible allocations in the given budget set. However, for the purpose of this paper it is critical that the experimental paradigm allows conveying aggregate information about others' behavior. In this study's game, subjects are shown percentages, which are easily understood. In the graphical interface of Fisman et al. similar information would be much harder to convey.

This study is the first to connect the literature on social preferences with the recent literature on social norms and morality (see for instance Krupka & Weber 2009, Bicchieri & Xiao 2009, Dal Bo & Dal Bo 2009). Both those strands of the literature are interested in explaining why people deviate from selfishness. But the two approaches are rather different. The literature on social preferences describes people's concern for others as an individual trait. There might be substantial heterogeneity between individuals; some people care for social welfare, some people care for equality, some do not care for others at all. But at least people know what matters to them. In a non-strategic environment like the MDG, players' preferences are supposed to be robust to different contexts and states of mind.[5] Consequently, according to this literature neither information nor self-reflection should have any effect on the revealed social preferences of the deciders.

Prediction 1: *Social preferences are individual traits. Neither information nor self-reflection will have an effect on people's choices in the MDG.*

[5] Charness & Rabin (2001) do think, however, that preferences might change in a strategic environment due to negative or positive reciprocity.

In a public good game Dal Bo & Dal Bo (2009) provide participants with messages that define moral behavior. They find that if subjects are told that full contribution is moral their cooperation rates increase but still fall over time. Whereas Dal Bo & Dal Bo tell subjects what is supposed to be moral, Engel and Kurschilgen (2012) ask subjects for their own normative standards and show that this can indeed sustain cooperation in a public good game at very high levels. Also in a public good game, Engel, Kube and Kurschilgen (2011) show that information radically changes behavior when the information is anti-social but barely when it is pro-social. While those experiments show that social behavior changes, they cannot show that social preferences change since strategic considerations and beliefs are paramount in public good games.

The dictator game eliminates strategic concerns. In the framework of a standard dictator game, Bicchieri & Xiao (2009) manipulate dictators' empirical expectations by telling them what the majority of participants did in previous studies and their normative expectations by telling them what the majority thought should be done. They find empirical expectations to have a much stronger effect than normative expectations. In contrast to this study, however, the information subjects receive in Bicchieri & Xiao is not representative but selective. Krupka & Weber (2009) test the effect of focusing and information on choices in a binary dictator game. In the focusing treatments they have subjects think about what others would do ("descriptive focus"), or think about what other others said one should do ("injuctive focus"). In the informational treatment, subjects observe the choices of four previous players. Their results show that both focusing and information increases pro-social behavior.

Once one leaves the conceptualization of social preferences as stable individual traits and accepts the idea that different contexts may entail different norms which in turn may change preferences, it can be useful to think of individuals as having a dual self. On the one hand, people are egoists who maximize their own material wealth. On the other hand, the very same people care intrinsically for abiding with norms. As the relevant norms vary from one context to the other, so does the self-perception of the individual. In one situation I might perceive myself as a player who legitimately maximizes her own profit whereas in another situation the very same behavior would cause me moral remorse. An early advocate of this idea is Adam Smith (1790) who claims that humans have both "passive feelings" that are "selfish" and "active principles" that are "noble" and "generous". He further reasons that if we act

according to those noble principles we "attain […] satisfaction" not only from "praise but [from] praise-worthiness" (Smith 1790: III.ii.32). To experience that satisfaction "we must become the impartial spectators of our own character and conduct".

Akerlof & Kranton (2000 and 2010) introduce the concept of identity utility to capture such contextual changes of preferences. We can think of the following simple utility function:

$$U = \pi(a) - \gamma\, D(a, \tilde{a}) \qquad (5)$$

where a represents the actual behavior of an individual and $\tilde{a}$ is the norm. Utility increases with material payoffs $\pi(a)$ and decreases with every deviation from the norm $D(a, \tilde{a})$. How much a norm deviation hurts depends on the weight γ, which can be interpreted as the importance of a norm. According to this model, a person would only deviate from material selfishness if she believes there is a certain norm $\tilde{a}$ that does not coincide with profit maximization. In this case, the utility maximizing behavior represents a compromise between profit maximization and norm compliance. Consequently, information and self-reflection might have an effect on behavior if they change either the norm $\tilde{a}$ or its importance γ.

However, Akerlof & Kranton lack some precision as to what exactly constitutes a norm. Their *informal* conceptualizations of a norm allude to two distinct sources: The first source is external. People watch how others behave and deduce the norm from what most people do. I call this a norm of conformity. On the other hand, the source of a norm can also be internal. People look at their own moral standards and deduce the norm from what they personally believe to be morally right. I call this a norm of morality. Conformity requires information about others' behavior. Morality requires self-reflection about one's own moral standards. This is the rationale for the two experimental treatments.

In the *Info* treatment, deciders are informed about the choices of previous deciders. Specifically, they see that the absolute majority of previous participants chose Option A in each of the 36 decision tasks.[6] Consequently if information reinforces conformity, choices in *Info* should be both more selfish and more homogenous than in the *Baseline*.

[6] Only the size of the majority varies from 53% to 99%. See Appendix C.

Prediction 2: *Social preferences are subject to conformity. Information will make choices in the MDG more selfish and more homogeneous.*

In the *Reflect* treatment, subjects are asked to state their moral judgments. Using those as proxies for $\tilde{a}$ and following Akerlof & Kranton, actual choices should be between the selfish optimum and deciders' moral judgments. The larger γ the closer a will be to $\tilde{a}$. Consequently if self-reflection enhances morality, choices in the *Reflect* treatment should be closer to $\tilde{a}$ and thus further away from the selfish optimum than in the *Baseline*. Previous experiments on social preferences suggest that most people who deviate from material selfishness are guided by either equality or social welfare concerns (see for instance Iriberri & Rey 2010, Charness & Rabin 2001, Fisman et al. 2007). When deciders are richer than recipients these two concerns trigger the same behavior. This implies that if self-reflection reinforces morality, when deciders are ahead we should expect more creating in *Reflect* compared to the *Baseline*. However when deciders are poorer than recipients, the concern for equality postulates exactly the opposite than the concern for social welfare. Consequently if self-reflection reinforces morality, when deciders are behind we should expect social welfare maximizers to create more and equality maximizers to destroy more in *Reflect* compared to the *Baseline*. This would lead to more heterogeneity of choices in *Reflect* than in the *Baseline*.

Prediction 3: *Social preferences are subject to morality. Self-reflection will make choices in the MDG less selfish and more heterogeneous.*

IV. Results

In the MDG, every deviation from the payoff maximizing Option A implies a cost of 10 tokens for the decider. On average[7], deciders in the baseline paid 37 tokens to deviate from material selfishness, compared to 32 tokens in *Info* (Mann-Whitney ranksum test, N=183 p=0.4330, two-sided), and 71 tokens in *Reflect* (ranksum test, N=159, p=0.0053, two-sided). Information has a totally different effect than self-reflection. The former seems to barely influence preferences whereas the latter pushes

[7] As mentioned before, I am interested in choices that are deliberate and conscious and not due to confusion or cognitive overload. Therefore, throughout the results section, I only consider deciders who choose consistent with GARP: 79% in the *Baseline*, 88% in *Info*, and 81% in *Reflect*. However, all main results also hold if one includes the inconsistent deciders.

behavior significantly away from material selfishness. As shown in the cumulative distribution of Figure 1, in the *Baseline* 50% of deciders are not willing to pay a single token but prefer to just maximize their material payoffs. This number increases slightly to 56% (ranksum test, N=183, p=0.4760, two-sided) for deciders who were informed about behavior in the baseline, and drops considerably to 33% (ranksum test, N=159, p=0.0705, two-sided) for deciders who had previously been asked about their moral judgment. On the other side of the distribution, only 4% of the *Baseline* deciders were willing to give away the maximum possible amount of 180 tokens. This number even drops slightly to 2% in *Info* (ranksum test, N=183, p=0.3532, two-sided) and jumps to 15% in *Reflect* (ranksum test, N=159, p=0.0168, two-sided).

Result 1: *Self-reflection makes deciders deviate substantially more from the selfish optimum. Information has no effect.*

Figure 1: Mean Choices across Treatments
Note: Cumulative distribution function, displaying the share of individuals willing to pay (up to) a certain price to deviate from material selfishness. Every deviation from the selfish optimum implied forgone payoffs of 10 tokens.

In which situations are deciders willing to give up own payoffs? When they are richer than the recipients, many deciders forego higher own payoffs in order to increase the income of the recipient. In the *Baseline*, 40% of deciders do so. This number decreases slightly to 35% in *Info* (ranksum test, N=183, p=0.5029, two-sided) and jumps to 64% in *Reflect* (ranksum test, N=159, p=0.0089, two-sided). When deciders are poorer than recipients, 32% of the *Baseline* deciders choose to spend money, compared to 35% (ranksum test, N=183, p=0.7420, two-sided) in *Info* and 54% in *Reflect* (ranksum test, N=159, p=0.0172, two-sided). Interestingly however, self-reflection not only makes more people spend money on creating income but also on destroying income: 10% of deciders destroy recipients' income in *Reflect*, compared to only 4% (ranksum test, N=159, p=0.1541, two-sided) in the *Baseline* and 0% in *Info* (ranksum test, N=102, p=0.0099, two-sided). People who create when ahead and destroy when behind believe in the norm of equality whereas people who create both when ahead and when behind have social welfare as their normative goal. Apparently, self-reflection reinforces of one's privately held norms, no matter whether the specific norm is equality or social welfare maximization. When these norms are in conflict to one another, as it is the case when the deciders are behind, self-reflection not only leads to a stronger deviation from material selfishness but also to more heterogeneity of behavior (variance ratio test, N=159, p=0.0190).

Result 2: *Self-reflection makes preferences more heterogeneous as it reinforces deciders' home-grown normative concerns, be it social welfare or equality. Information has no effect.*

In fact, deciders' moral judgments are even more extreme than their actual behavior. As predicted by the concept of identity utility, behavior in the *Reflect* treatment is a compromise between one's moral judgment and payoff maximization. Figure 2 illustrates this. Every dot represents the choice behavior of one individual decider. They are sorted by their cutoff ρ and σ (see Table 2). A positive/negative ρ (σ) represents creating/destroying when ahead (behind). Hence, social welfare maximizers are located in the northeast quadrant, equality maximizers in the southeast quadrant, and payoff maximizers in the origin. Identity utility predicts that independent of the content of the norm, actual behavior will be closer to selfish payoff maximization than one's normative ideal. The data clearly support this. Of the 21 deciders who by their judgments in Panel A of Figure 2 can be identified as social welfare maximizers ($\rho \geq 0$, $\sigma > 0$), 13 move closer to the origin once money is at stake (in Panel B) and only one individual moves further away.

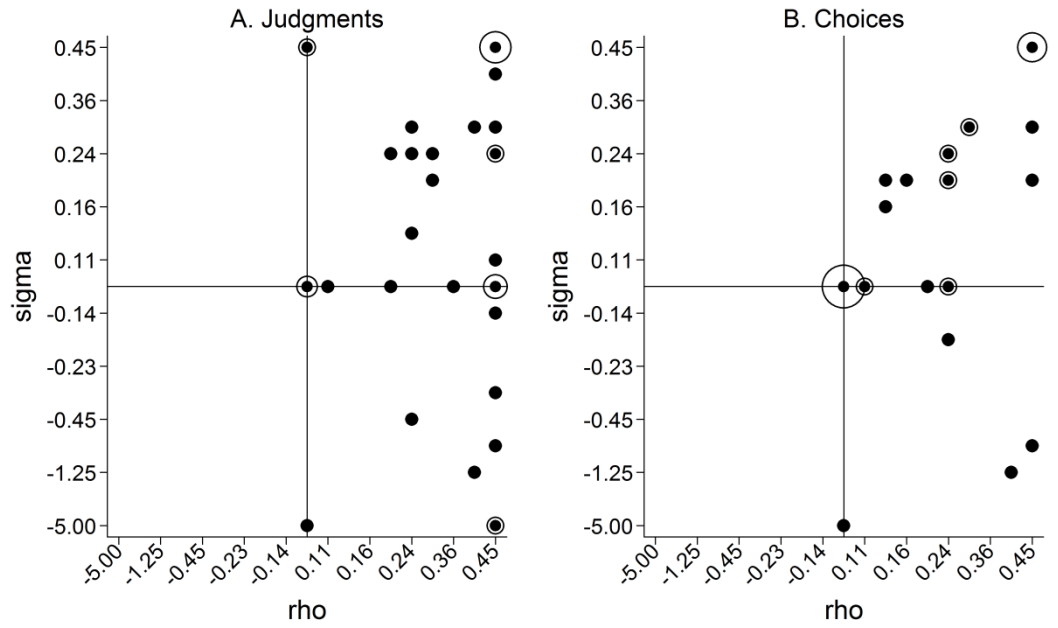


Figure 2: Individual Judgments and Choices in *Reflect* Treatment

Note: Every dot denotes an individual decider. The circles around the dots indicate that more than one decider had the same choice pattern. The larger the circles, the more deciders. The scales of the axes follow the logic of Table 2, i.e. that destroying 84 and creating 84 are equivalent deviations from payoff maximization. They are thus plotted at the same distance from the origin. They are then labeled by their cutoff ρ and σ, i.e. -5.00 for destroying and 0.45 for creating.

In total, the actual price deciders pay for complying with the norm is significantly smaller than the price they judge to be morally right (Wilcoxon signed-rank test, N=21, p=0.0010, two-sided). Also the eight inequality averse deciders ($\rho \geq 0, \sigma < 0$) have moral judgments that are more extreme than their actual choices (signed-rank test, N=8, p=0.0287, two-sided). This case is especially interesting since those people appear to judge it morally right to destroy even more of the other person's income than what they actually do once it is costly. Finally, the seven people who comply with both welfare maximization and inequality aversion as they have $\rho > 0$ and $\sigma = 0$ also move closer to the origin in their choices (signed-rank test, N=7, p=0.0311, two-sided). In contrast, the three individuals who believe payoff maximization to be morally right ($\rho = 0, \sigma = 0$) stick to this when money is at stake.

Of the 13 deciders who act like payoff maximizers in the MDG, five individuals believe in social welfare maximization, three in equality, and two in both.

Result 3: *Revealed preferences in the Reflect treatment represent a compromise between deciders' moral judgments and the selfish optimum, independent of whether people are concerned with social welfare or equality.*

So far, this study has shown that social preferences are very robust to information about the behavior of other people but change substantially when people are induced to make a moral judgment prior to their choice. The experimental data contradicts Predictions 1 and 2 and supports Prediction 3, which suggests that morality has a much stronger effect on social preferences than conformity.

However, in reality information might be available before making a moral judgment or between my judgment and my choice. How robust are people's moral judgments to information? How much do actual choices deviate from moral judgments once information kicks in? To test these questions, I run two additional treatments.

The *Info+Reflect* treatment combines the two elements of *Info* and *Reflect*. After reading the instructions, subjects are informed that the game has been run before with more than 100 deciders. They are then asked to make their moral judgments while seeing on their screens for each of the 36 tasks which percentage of deciders in the *Baseline* preferred Option A. When playing the MDG they are reminded on their screens for each of the 36 tasks of both their own moral judgments and which percentage of deciders in the *Baseline* preferred Option A. 96 new participants (48 deciders) took part in this treatment.

In the *Reflect+Info* treatment the order of the two elements is reversed. First, subjects are asked to state what they believe to be "morally right", just as in the *Reflect* treatment. Thereupon they are informed that the game has been run before with more than 100 deciders. When playing the MDG they are reminded on their screens for each of the 36 tasks of both their own moral judgments and which percentage of deciders in the *Baseline* preferred Option A. 96 new participants (48 deciders) took part in this treatment.

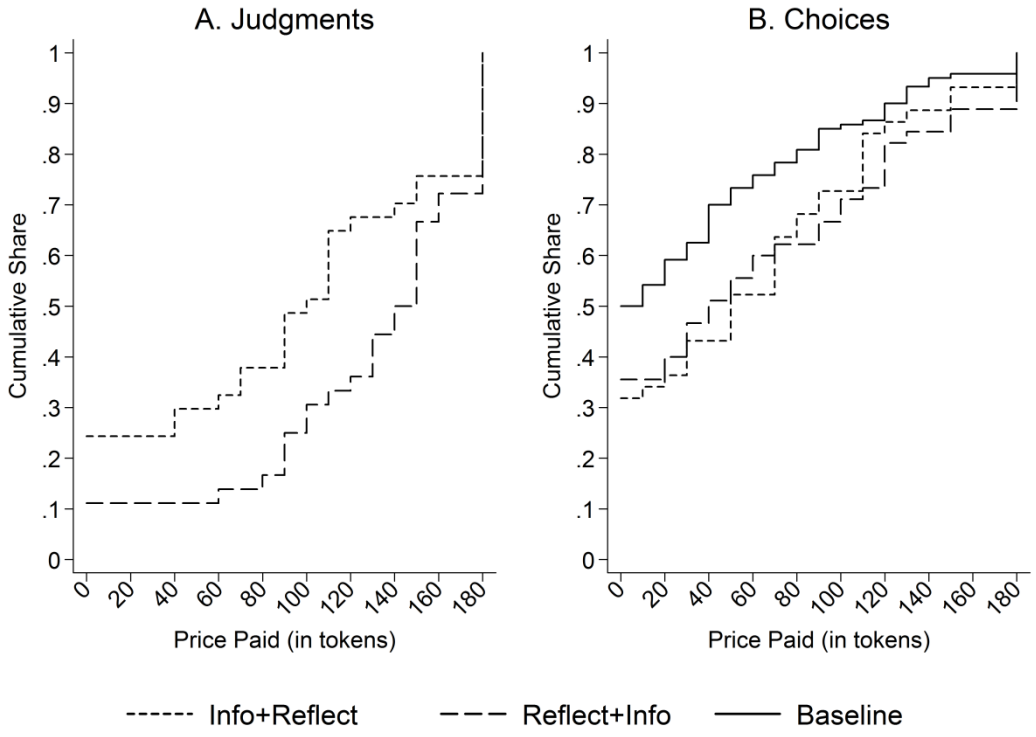


Figure 3: Mean Judgments and Choices across Treatments

Note: Cumulative distribution function, displaying the share of individuals willing to pay (up to) a certain price to deviate from material selfishness. Every deviation from the selfish optimum implied forgone payoffs of 10 tokens.

Panel A of Figure 3 shows the first main result from the additional treatments[8]. If subjects are informed about others' behavior beforehand, their moral judgments change substantially. In *Info+Reflect* the average decider believes one should only spend 95 tokens to comply with one's norm, compared to 129 tokens in *Reflect+Info* (ranksum test, N=89, p=0.0484, two-sided) and 130 tokens in *Reflect* (ranksum test, N=83, p=0.0301, two-sided). Strikingly however, the difference in moral judgments does not translate into actual choices. As shown in Panel B of Figure 3, the

[8] Also with the additional treatments, we restrict the analysis to the GARP consistent deciders, 92% in the *Info+Reflect* treatment and 94% in *Reflect+Info*. However, the main results also hold if we include the non-consistent, too.

willingness to spend money in the MDG is virtually identical in the two treatments. The average price paid is 60 tokens in *Info+Reflect* and 62 tokens in *Reflect+Info* (ranksum test, N=89, p=0.9900, two-sided). In fact, neither *Info+Reflect* (ranksum test, N=83, p=0.5548, two-sided) nor *Reflect+Info* (ranksum test, N=84, p=0.5694, two-sided) are significantly different from *Reflect*. In contrast, both *Info+Reflect* (ranksum test, N=164, p=0.0125, two-sided) and *Reflect+Info* (ranksum test, N=165, p=0.0217, two-sided) are significantly different from the *Baseline*.

Result 4: *Combined with self-reflection, information has a strong effect on people's moral judgments but no effect on their actual choices.*

In the framework of Akerlof & Kranton (2000 and 2010), the distance between a decider's moral judgment and her actual behavior indicates how much value she attaches to norm compliance. The moral judgments that were stated by the informed deciders appear to be more robust than those of the uninformed deciders in the sense that subjects deviate less from them when they play the payoff-relevant MDG. The comparison of the treatments *Reflect*, *Reflect+Info*, and *Info+Reflect* suggests that a deviation from one's normative ideal causes more disutility if the norm was grounded on information. As a consequence, people deviate less. This in turn happens to cancel out the effect of a lower normative standard in the first place.

Result 5: *Deciders deviate less from their moral judgments when those were grounded on information.*

V. Conclusion

This paper has tested experimentally the effect of information and self-reflection on people's social preferences in a modified dictator game (MDG). According to the literature on social preferences, these are stable traits of individuals that should not be altered by either information or self-reflection. In fact, this study finds that information has no effect at all on people's preferences. In contrast, self-reflection changes preferences substantially. Whilst information has no effect on people's choices, it does influence their moral judgments. Moreover, deciders whose moral judgments were stated knowing how other people behaved deviate much less from their judgments in the MDG. This suggests that people feel more strongly bound to judgments when those are well-informed.

The fact that information has virtually no effect on subjects' choices makes the strong effect of self-reflection even more striking. The way the information is presented in the *Info* treatment, subjects see that in every single decision task the absolute majority of previous participants chose Option A. Still, this apparently does not induce people to follow the crowd. In contrast, in the *Reflect* treatment subjects know at the moment they are stating their moral judgments that with 50% probability they will be deciders in the MDG. Hence, they could easily state more selfish moral judgments. But they do not. At the moment they play the MDG deciders know their role. So they could easily neglect what they stated behind the veil of role uncertainty. But they do not. Deciders obviously feel bound by their previous statement. And still, they do not follow their moral judgment blindly but recognize the trade-off between norm compliance and their material self-interests.

These results are very much in line with the idea of identity utility proposed by Akerlof & Kranton (2000). Moreover, they yield interesting new insights to the relationship between social preferences and the Akerlof & Kranton model. First, the source of the norm matters. Deciders' social preferences in this experiment were extremely robust to conformity but highly susceptible to morality. Second, the importance of morality holds independent of the specific normative concern of the individual deciders. This leads to preferences becoming more heterogeneous as self-reflection reinforces people's home-grown normative concerns. Some people believe in social welfare, others in equality. When deciders are poorer than recipients these two normative goals clash. Self-reflection makes social welfare maximizers create more and equality maximizers destroy more. And their moral judgments are even further apart than their actions. This might have interesting implications in strategic settings. A follow up study could investigate whether self-reflection increases the likelihood of conflict in situations where people have divergent home-grown normative concerns.

References

AFRIAT, SIDNEY (1967). "The construction of a utility function from expenditure data." International Economic Review **8**: 67-77.

AKERLOF, GEORGE A. and RACHEL KRANTON (2000). "Economics and Identity." Quarterly Journal of Economics **115**: 715-753.

AKERLOF, GEORGE A. and RACHEL KRANTON (2010). Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being. New Jersey: Princeton University Press.

ANDREONI, JAMES and JOHN MILLER (2002). "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." Econometrica **70**: 737-753.

BICCHIERI, CHRISTINA and ERTE XIAO (2009). "Do the Right Thing: But Only if Others Do So." Journal of Behavioral Decision Making **22**: 191-208.

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

CHARNESS, GARY and MATTHEW RABIN (2002). "Understanding Social Preferences with Simple Tests." Quarterly Journal of Economics **117**: 817-869.

DAL BO, ERNESTO and PEDRO DAL BO (2009). "The Effects of Moral Suasion on Cooperation." NBER Working Paper Series.

ENGEL, CHRISTOPH and MICHAEL KURSCHILGEN (2012). "The 'Jurisdiction Of The Man Within' Intrinsic Norms in a Public Goods Experiment." Working Paper.

ENGEL, CHRISTOPH, SEBASTIAN KUBE and MICHAEL KURSCHILGEN (2011). "Can we manage first impressions in cooperation problems? An experimental study on "Broken (and Fixed) Windows." Max Planck Institute for Research on Collective Goods Preprints 2011/05.

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.

FISMAN, RAYMOND, SHACHAR KARIV and DANIEL MARKOVITS (2007). "Individual Preferences for Giving." American Economic Review **97**: 1858–1876.

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

IRIBERRI, NAGORE and PEDRO REY-BIEL (2010). "Elicited Beliefs and Social Information in Modified Dictator Games: What Do Dictators Believe Other Dictators Do?." Working Paper.

IRIBERRI, NAGORE and PEDRO REY-BIEL (2011). "The role of role uncertainty in modified dictator games." Experimental Economics **14**: 160-180.

KRUPKA, ERIN and ROBERTO A. WEBER (2009). "The Focusing and Informational Effects of Norms on Pro-Social Behavior." Journal of Economic Psychology **30**: 307-320.

SMITH, ADAM (1790). The Theory of Moral Sentiments, or, An Essay Towards an Analysis of the Principles by which Men Naturally Judge Concerning the Conduct and Character, First of Their Neighbours, and Afterwards of Themselves. To which is Added, a Dissertation on the Origin of Languages. London, Strahan.

VARIAN, HAL R. (1982). "The Nonparametric Approach to Demand Analysis." Econometrica **50**: 945-972.

Appendix A: Experimental Instructions

General Information

Welcome to our experiment!

If you read the following explanations carefully, you will be able to earn a substantial sum of money, depending on the decisions you make. It is therefore crucial that you read these explanations carefully.

During the experiment there shall be absolutely no communication between participants. Any violation of this rule means you will be excluded from the experiment and from any payments. If you have any questions, please raise your hand. We will then come over to you.

During the experiment we will not calculate in euro, but instead in tokens. Your total income is therefore initially calculated in tokens. The total number of tokens you accumulate in the course of the experiment will be transferred into Euro at the end, at a rate of

100 tokens = 1 Euro.

At the end you will receive from us the **cash** sum, in euro, based on the number of tokens you have earned.

The Experiment

In the experiment, there are two roles: **decider** and **recipient**.

At the beginning of the experiment you will be randomly allotted one of the two roles. One half of the participants will be deciders, the other half will be recipients. During the entire experiment, you will remain in the same role.

On your computer screen you will be shown **4 tables**, one after the other. Every table consists of **9 decision tasks**.

A decision task could for example read as follows:

	Option A		Option B	
Decider (You)	12		10	
Recipient	5		7	Your decision (A or B):

In every decision task the decider has to choose between **Option A** and **Option B**. The two options define how many **tokens** the decider gets and how many the recipient gets.

In this example the decider gets 12 tokens and the recipient 5 tokens if the decider chooses Option A. If the decider chooses Option B, the decider gets 10 tokens and the recipient 7 tokens.

In every decision task the computer will **randomly** match every decider with a different recipient. Thus the decider-recipient pairs change in every decision task.

The decider will never know the identity of the recipient.
The recipient will never know the identity of the decider.

At the end of every table please press the "OK" button on the lower right hand side of your screen. Only after pressing "OK" your decisions are saved and become effective. You will then be shown the next table.

Payoffs

At the end of experiment the computer will **randomly** pick one decision task out of every table. The computer thus picks in total **4 decision tasks**, one from every table. The corresponding token amounts from those 4 decision tasks will be added and changed into Euros.

If you are **decider**, your payoffs only depend on your own choices and on the random draw at the end of the experiment.

If you are **recipient**, your payoffs only depend on the choices of the corresponding decider and the random draw at the end of the experiment.

Appendix B: Additional screen in *Info* treatment

This Experiment has been run before with more than 100 Deciders.
In the column on the right hand side of your screen you can see how the Deciders in those previous Experiments decided. Specifically, you will be shown which percentage of Deciders chose Option A or Option B in the corresponding Choice Task.

Appendix C: Information about choices of previous players (by panel)

	Ahead Create	**Ahead Destroy**	**Behind Create**	**Behind Destroy**
1	89% chose A	95% chose A	91% chose A	95% chose A
2	89% chose A	95% chose A	92% chose A	91% chose A
3	87% chose A	97% chose A	89% chose A	95% chose A
4	83% chose A	98% chose A	88% chose A	90% chose A
5	76% chose A	99% chose A	84% chose A	90% chose A
6	68% chose A	97% chose A	76% chose A	89% chose A
7	64% chose A	97% chose A	74% chose A	91% chose A
8	58% chose A	97% chose A	68% chose A	90% chose A
9	53% chose A	97% chose A	67% chose A	89% chose A

Appendix D: Additional screen in *Reflect* treatment

Before the computer randomly determines who will be Decider and who will be Recipient, we would like to know your opinion.
We would like to know from you:
Which of the two Options (A or B) do you find morally right?
The answers to these questions will be kept anonymous. No other participant will get to know them at any time. Your answers to these questions are not relevant for your payoffs.

Conclusion

This dissertation has examined how norms guide our behavior in the face of uncertainty. Norms can have many origins: institutionally imposed by a legislator, behavioral patterns within a social group, but also moral conceptions inherent to the individual. Equally, there are also many potential sources of uncertainty: nature, other people's behavior, and even one's own preferences. I have tackled several sub-questions of this overarching research agenda in four separate Chapters.

The first Chapter has compared two institutional arrangements for the market of copyrights. This market is characterized by the highly skewed distribution of earnings and their unpredictability. In the first institutional setting, the copyright must be traded behind the veil of uncertainty for a fixed fee. In the unlikely, but possible event of high success in the market, the licence fee is nonetheless binding for the two parties. The second setting introduces renegotiation in the shadow of legal intervention. In case the work turns out a bestseller, the artist may appeal to a third party who is entitled to adjust the fee. This situation corresponds to a provision from German copyright law, the so-called "bestseller paragraph". In such a market, conflicting fairness norms between buyers and sellers are amongst the biggest obstacles to trade. Yet even when the parties have reached an agreement in the first place ("ex ante"), substantial discontent may arise as soon as the true value of the commodity is revealed ("ex post").

The experimental results suggest that, in the presence of the provision, copyrights trade at lower prices. Second, they trade more often as the acceptance level of sellers is reduced more than the willingness to pay of buyers. Higher acceptance is a strong indicator of enhanced ex-ante fairness. Moreover, the provision reduces perceived ex-post unfairness for buyers. Rather surprisingly, though, there is not a similar effect for sellers. In the provision treatment, a third party, the so-called umpire, was asked to determine her "appropriate purchase price" in case the commodity had a high value. Even though the experimental design was rather prone to highlight ex-ante equality, ex-post equality turns out the umpires' single distinct fairness norm.

The Chapter has used the movie industry as the primary example because there is precise data on this market. Yet, given the neutral frame of the experimental design,

the insights should also be relevant to other copyright markets, characterised by comparable unpredictability of earnings, like exhibitions and music. Other markets with highly skewed earnings, like venture capital, might also be affected by a similar fairness problem. Future work might want to introduce different allocations of market power and test whether under such less favourable conditions the welfare enhancing effect of the provision is substantially reduced. In addition, one could make the sellers' production of the copyright endogenous, to study the dynamic effects of different institutional arrangements. Another extension could allow the buyers' effort to influence the probability of success. One might also want to study to which degree reputation is a substitute for legal intervention.

Chapter 2 has studied the effect of customary law on cooperation. The experimental results from a standard linear public goods game suggest that customary law helps participants overcome a social dilemma. If there are no sanctions, the effect basically coincides with the behavioural effect of what public international law calls comity. In essence, customary law governs behaviour since normative expectations and behaviour coevolve. Whether the rule invokes the authority of the law is at best irrelevant, if not detrimental. Yet the authority of the law becomes instrumental as soon as there are sanctions. If combined with comity, sanctions crowd out some of the beneficial effect. If the rule originates in law, however, the authority of the law and the threat of sanctions reinforce each other.

An obvious next step for studying the coevolution of behaviour and normative expectations would be to qualitatively explore normativity in international relations. In one respect, the character of states as corporate actors makes this even easier than with individuals. While the formation of the individual will occurs in her *forum internum*, the formation of the corporate will is open to public scrutiny. Through freedom of information legislation, even internal government deliberation is made accessible. It seems plausible that normativity in international relations is not fundamentally different from normativity in personal relations.

Chapter 3 has explored whether cooperative behavior reacts to selective pre-play information about other, unrelated groups. It uses an experimental framework that is sufficiently rich, so that several potentially conflicting behavioral norms might emerge. The data show that the aggregate level of cooperativeness is dramatically reduced by giving subjects examples of uncooperative behavior. The opposite intervention is less powerful, but still effective. If participants in an otherwise

uncooperative environment see examples of successful groups, cooperation rates are stabilized, while they quickly decayed otherwise. A likely reason for these findings is that cooperation is path-dependent, i.e., long-run behavior is strongly correlated with initial contribution rates – which, in turn, are a reaction to the pre-play information given to the subjects. The latter might be due to a change in subjects' expectations about others' behavior.

These findings have a number of important implications. First, they clearly point to the relevance of pre-game communication – a factor which has only lately started to receive significant attention in the literature. While the existing literature usually focuses on self-chosen cheap talk messages, this Chapter has demonstrated that also exogenously selected, one-way information about other players can alter how players act in subsequent games. In particular the findings of the second series of experiments might be of interest to this literature. They suggest a possible channel through which the observed effects of cheap talk are mediated, namely through the alteration of subjects' pre-game expectations.

Second, and closely related, is the finding that there seems to be a strong asymmetry in the strength of the reaction to the mechanism. By giving subjects examples of uncooperative behavior, the aggregate level of cooperativeness is immediately reduced. By contrast, cooperative examples need time to unfold their beneficial effect. Bad impressions carry more weight than good ones, so that creating bad behavior is seemingly easier than producing good behavior. Thus, one might expect that also other mechanisms which build on the manipulation of beliefs perform differently, depending on the content of the experience as well as on the specific environment in which they occur.

Taken together, the results of Chapter 3 underline the power and importance of information and experience in shaping cooperative behavior. The bottom line is that observation matters. Interestingly, people do not only learn from what they experience themselves. They also seem to learn "vicariously", by observing others, or by seeing the results. The effect is even present when participants are told that the information they are receiving is selective. Subjects' behavior critically depends on pre-game expectations, which is shown to be easy to deteriorate in a complex setting – simply by providing subjects the opportunity of vicarious learning.

Managing first impressions might certainly be less effective if the large majority of addressees know better. The intervention requires a sufficient degree of uncertainty. Yet in political reality, quite a few public goods are characterized by deep conceptual and factual uncertainty. Problems like climate change are heavily contested among scientists and not well understood by many. Addressees have to trust expertise. If in the eyes of addressees the underlying social problem is opaque, they are also likely to be uncertain how others will react to it. Most importantly, addressees face behavioral uncertainty whenever they newly enter a community. They do not know local norms, nor do they know how determined the group is when it comes to enforcing them.

Future research could try to explore the interaction between such simple and other, more complex and intrusive mechanisms. Another important, though challenging, next step would certainly be to move into the field and to investigate the effectiveness of means to foster cooperation in a more natural setup.

The last Chapter of this dissertation has combined elements of Chapter 2, where subjects are asked for their own normative standards, and Chapter 3, where subjects are briefed about the behavior of previous participants. While those experiments show that social behavior changes, they cannot show that preferences change since strategic considerations and beliefs are paramount in public good games. Moreover, whilst Chapters 2 and 3 have looked at aggregate behavior of all experimental participants, the last Chapter has introduced an additional layer of analysis by distinguishing between different player "types".

Chapter 4 has studied the effect of information and self-reflection on people's social preferences in a modified dictator game (MDG). According to the literature on social preferences, these are stable traits of individuals that should not be altered by either information or self-reflection. This Chapter suggests that information has indeed no effect at all on people's preferences. In contrast, self-reflection changes preferences substantially. Whilst information has no effect on people's choices, it does influence their moral judgments. Moreover, deciders whose moral judgments were stated knowing how other people behaved deviate much less from their judgments in the MDG. This suggests that people feel more strongly bound to judgments when those are well-informed.

The fact that information has virtually no effect on subjects' choices makes the strong effect of self-reflection even more striking. The way the information is presented in

the *Info* treatment, subjects see that in every single decision task the absolute majority of previous participants chose Option A. Still, this apparently does not induce people to follow the crowd. In contrast, in the *Reflect* treatment subjects know at the moment they are stating their moral judgments that with 50% probability they will be deciders in the MDG. Hence, they could easily state more selfish moral judgments. But they do not. At the moment they play the MDG deciders know their role. So they could easily neglect what they stated behind the veil of role uncertainty. But they do not. Deciders obviously feel bound by their previous statement. And still, they do not follow their moral judgment blindly but recognize the trade-off between norm compliance and their material self-interests.

These results are very much in line with the idea of identity utility. Moreover, they yield interesting new insights to the relationship between identity and social preferences. First, the source of the norm matters. Deciders' social preferences in this experiment were extremely robust to conformity but highly susceptible to morality. Second, the importance of morality holds independent of the specific normative concern of the individual deciders. This leads to preferences becoming more heterogeneous as self-reflection reinforces people's home-grown normative concerns. Some people believe in social welfare, others in equality. When deciders are poorer than recipients these two normative goals clash. Self-reflection makes social welfare maximizers create more and equality maximizers destroy more. And their moral judgments are even further apart than their actions. This might have interesting implications in strategic settings. A follow up study could investigate whether self-reflection increases the likelihood of conflict in situations where people have divergent home-grown normative concerns.

Deutsche Zusammenfassung

Das Leben steckt voller Ungewissheiten. Bei den meisten Entscheidungen, die wir zu treffen haben, kennen wir die genauen Folgen unseres Handelns nicht. Wenn wir vor einer roten Ampel stehen, wissen wir zum Beispiel häufig nicht, wie gefährlich die entsprechende Kreuzung ist oder ob die Polizei einen versteckten Blitzer aufgestellt haben könnte. Wenn wir uns entschließen, ein Haus zu kaufen, wissen wir nicht, ob wir uns mit unseren Nachbarn verstehen werden und wie sich der Wert des Hauses über die Jahre entwickeln wird, falls wir es wieder verkaufen wollen. Wenn wir überlegen, einer Hilfsorganisation Geld zu spenden, können wir nie ganz sicher sein, dass das Geld bei den Betroffenen ankommt und ob die Hilfsmaßnahmen tatsächlich die Lebensbedingungen der Menschen verbessern.

In dieser ungewissen Welt helfen uns Normen dabei, Entscheidungen zu treffen, indem Sie uns Orientierung geben. Normen können unterschiedlichen Ursprungs sein. Sie können Gesetze sein, typische Verhaltensmuster einer sozialen Gruppe oder persönliche moralische Grundsätze. Wir überfahren keine roten Ampeln, weil das gegen das Gesetz verstößt; völlig unabhängig von der speziellen Situation an einer bestimmten Kreuzung. Wir bauen ein Haus, weil die meisten unserer Freunde und Kollegen Häuser besitzen. Und wir spenden, weil wir glauben, dass irgendetwas getan werden muss.

In vier eigenständigen Kapiteln erforsche ich, wie Normen unser Verhalten beeinflussen. Dabei wende ich experimentalökonomische Methoden auf Fragestellungen an, die sich an der Schnittstelle zur Rechtswissenschaft (insb. Kapitel 1 und 2), Soziologie (insb. Kapitel 3) und Sozialpsychologie (insb. Kapitel 4) befinden.

Das erste Kapitel untersucht den Markt für Urheberrechte. Stellen Sie sich folgende Situation vor: Ein Autor hat ein Drehbuch geschrieben und ein Produzent möchte das Recht erwerben, daraus einen Film zu machen. Zum Zeitpunkt, da beide Parteien sich einigen müssen, weiß keine der beiden, wieviel der Film einspielen wird. Erfolg an der Kinokasse ist sehr ungleich verteilt: Wenige Filme erzielen einen enormen Gewinn, während die große Mehrzahl kaum die Produktionskosten einzupielen vermag. Es gibt keine zuverlässigen Methoden, Kassenerfolg vorherzusagen. In so

einer Situation ist es für beide Parteien schwierig, ihren Reservationspreis festzulegen, geschweige denn, sich auf ein für beide Seiten akzeptables Geschäft zu einigen. Der deutsche Gesetzgeber glaubt, dass diese Situation einen regulatorischen Eingriff erfordert. Das ist im sogenannten „Bestsellerparagraphen" des Urheberrechts festgelegt. Dieser besagt, dass der Urheber das Recht hat, im Nachhinein eine Änderung des Vertrags zu verlangen, sollte die ursprüngliche Vereinbarung „in einem auffälligen Missverhältnis zu den Erträgen [...] aus der Nutzung des Werkes stehen" (§32a I 1 UrhG).

Zusammen mit Christoph Engel entwerfe ich ein Experiment, das die wichtigsten Eigenschaften des Marktes für Urheberrechte enthält. Eine Experimentalbedingung stellt den Markt *ohne* und eine andere *mit* dem Bestsellerparagraphen dar. Unter standardökonomischen Annahmen ändert die Vorschrift nichts am Gleichgewicht. Unter verhaltensökonomischen Annahmen jedoch führt die Vorschrift zu niedrigeren Preisen. Sie hat aber keinen Effekt auf die Anzahl der geschlossenen Verträge. Die Ergebnisse des Experiments suggerieren, dass die Vorschrift in der Tat zu niedrigeren Preisen führt und sich die Probanden fairer behandelt fühlen. Sowohl ex ante, als auch ex post.

Während sich das erste Kapitel mit kodifiziertem Recht befasst, beschäftigt sich das zweite Kapitel mit Gewohnheitsrecht. In den meisten Ländern der Welt ist Gewohnheitsrecht eine anerkannte Rechtsquelle. Es entsteht durch das Verhalten und den Willen derjenigen, die sich diesem Recht unterwerfen. Klassische Beispiele sind internationales Recht, die *lex mercatoria*, Handelsbräuche und Verhaltensregeln. Die typischen Bereiche des Gewohnheitsrechts zeichnen sich dadurch aus, dass hohe Ungewissheit darüber herrscht, wie sich die anderen Akteure verhalten werden, und, ob eine bestimmte Praxis bereits hinreichend etabliert ist, um Bindungswirkung zu entfalten.

Zusammen mit Christoph Engel untersuche ich die Wirkung von Gewohnheitsrecht in einem experimentellen Gemeinwohlspiel. Unsere Ergebnisse zeigen, dass Normen wirken, indem sie den beteiligten Akteuren Orientierung bieten. Die meisten Akteure halten sich die meiste Zeit an die Normen, die in ihrem Kontext gültig sind. Und sie erwarten, dass sich die anderen auch daran halten. Die Experimentalbedingung mit Gewohnheitsrecht verzeichnet wesentlich höhere Beiträge zum Gemeinwohl als die Kontrollbedingung.

Kapitel 3 verlässt den Bereich der Rechtsnormen und wendet sich sozialen Normen zu. Soziale Normen sind essenziell, wenn Menschen kooperieren müssen, um einen größeren sozialen Nutzen zu erzielen. Hinter vielen alltäglichen Situationen, wie zum Beispiel Umweltschutz, Landesverteidigung oder politischem Engagement, verbergen sich Kooperationsprobleme. In all diesen Fällen sehen sich die Akteure einem sozialen Dilemma ausgesetzt. Als Gemeinschaft geht es ihnen am besten, wenn jeder Einzelne seinen fairen Beitrag leistet. Aber individuell ist jeder in Versuchung, auf den Anstrengungen der anderen Trittbrett zu fahren und selbst nichts beizutragen. Frühere experimentelle Studien haben gezeigt, dass die meisten Gruppen es nicht schaffen, über einen längeren Zeitraum zu kooperieren.

Zusammen mit Christoph Engel und Sebastian Kube untersuche ich, welchen Effekt es hat, wenn man den Gruppen, bevor sie interagieren, Verhaltensbeispiele früherer Gruppen zeigt. In einer Experimentalbedingung erhalten die Teilnehmer Beispiele besonders kooperativer, in einer anderen, besonders unkooperativer Gruppen. Wir suggerieren somit den Teilnehmern eine mögliche (kooperative oder unkooperative) Verhaltensnorm. Das experimentelle Paradigma ist ein Gemeinwohlspiel mit Strafe und Revanche. Es ist hinreichend komplex, so dass sehr unterschiedliche Verhaltensnormen plausibel erscheinen.

Unsere Ergebnisse zeigen, dass die Beispiele einen sehr starken Effekt auf das Kooperationsverhalten der Gruppen haben. Zudem weisen sie auf eine interessante Asymmetrie hin. Die Teilnehmer reagieren wesentlich stärker auf die negativen als auf die positiven Beispiele. Eine zusätzliche Datenerhebung legt nahe, dass dies zur Ursache haben kann, dass optimistische Erwartungen über das Verhalten der Mitmenschen fragiler sind als pessimistische.

In den meisten Situationen des Alltags wissen wir, welche Norm relevant ist, und passen unser Verhalten automatisch und unbewusst an. Häufig jedoch finden wir neue Situationen vor und sind uns nicht sicher, wie wir uns *richtigerweise* verhalten sollen. Normen helfen uns, Orientierung zu finden. Man kann zwei Quellen unterscheiden, aus denen wir normative Orientierung beziehen. Extern, indem wir beobachten, wie sich andere Leute verhalten („Konformität“). Intern, indem wir uns fragen, was wir selbst als moralisch richtig beurteilen würden („Moralität“). Konformität benötigt Informationen über das Verhalten der anderen. Moralität bedarf der Selbstreflexion über die eigenen moralischen Grundsätze. Kapitel 4 untersucht den Effekt von Information und Selbstreflexion auf soziale Präferenzen.

Das Konzept der Selbstreflexion lässt sich auf Adam Smiths „Theory of Moral Sentiments“ zurückführen, der sich wiederum am religiösen Konzept des Gewissens orientierte. Das jüngere ökonomische Konzept des Identitätsnutzens erklärt die Veränderung von Präferenzen mittels einer Verschiebung der relevanten Normen; sei es unbewusst oder durch bewusste Selbstreflexion. Dies kontrastiert stark mit der heute weit verbreiteten Ansicht, dass Präferenzen stabile Merkmale eines Individuums sind. Die Literatur zu sozialen Präferenzen, zum Beispiel, akzeptiert zwar, dass unterschiedliche Menschen verschiedene Vorstellungen von Verteilungsgerechtigkeit haben, nicht jedoch, dass dieselben Personen von einer Situation zur anderen ihre Vorstellungen ändern.

Die Experimentalergebnisse legen nahe, dass Informationen keinen Einfluss auf soziale Präferenzen haben. Im Gegensatz dazu hat Selbsreflexion einen starken Effekt. Informationen über das Verhalten der anderen haben jedoch einen Effekt auf die Moralvorstellungen der Teilnehmer. Die Teilnehmer haben sehr heterogene Vorstellungen von Verteilungsgerechtigkeit. Während einige für Einkommensgleichheit eintreten, bevorzugen andere die Maximierung des Gemeinwohls. Selbstreflexion verstärkt das Gewicht der persönlichen Moralvorstellungen und führt somit zu mehr Heterogenität in den Präferenzen.

Ehrenwörtliche Erklärung

Mir ist die geltende Promotionsordnung der Wirtschaftswissenschaftlichen Fakultät der Friedrich-Schiller-Universität Jena bekannt. Ich habe die vorliegende Dissertation selbst angefertigt und keine Textabschnitte eines Dritten oder eigener Prüfungsarbeiten ohne Kennzeichnung übernommen. Alle von mir benutzten Hilfsmittel, persönliche Mitteilungen und Quellen habe ich an den entsprechenden Stellen kenntlich gemacht. Bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manusskriptes haben mich ausschließlich meine Koautoren unterstützt, d.h. Christoph Engel in Kapitel 1, 2 und 3, sowie Sebastian Kube in Kapitel 3. Den genauen Umfang habe ich in der Einleitung vermerkt. Ich habe nicht die Hilfe eines Promotionsberaters in Anspruch genommen. Dritte haben weder unmittelbar noch mittelbar geldwerte Leistungen von mir erhalten für Arbeiten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Ich habe diese Dissertation noch nicht für eine staatliche oder andere wissenschaftliche Prüfung eingereicht. Ich habe weder die vorliegende Dissertation noch eine in wesentlichen Teilen ähnliche Dissertation oder eine andere Abhandlung bei einer anderen Hochschule bzw. anderen Fakultät eingereicht.

Bonn, den 28.01.2013 .....