The impact of legal mechanisms on cooperation in infinitely repeated games: an experimental approach

a dissertation presented
by
Bethany Kirkpatrick
to
The Paris School of International Affairs

in partial fulfillment of the requirements
for the degree of
Master
in the subject of
International Economic Policy

Sciences Po, Paris
Paris, France
April 2014

Thesis advisor: Professor Emeric Henry Bethany Kirkpatrick

The impact of legal mechanisms on cooperation in infinitely repeated games: an experimental approach

Abstract

This paper presents experimental evidence on the impact of formal institutions on cooperative behaviour in infinitely repeated games. We examine the extent to which formal rules can prevent or limit opportunistic behaviour and promote cooperation when the informal punishment and reward schemes of repeated interactions cannot be supported in equilibrium. In particular, we study whether legal rules have extra-deterrent effects beyond altering the relative payoff to cooperation. Furthermore, we test for the presence and direction of spillover effects from environments covered by legal rules to those that are not.

We study the effect of legal rules on levels of cooperation in environments where, in the absence of formal sanctions, cooperation is otherwise not supported in equilibrium. We then compare play in environments where legal rules are explicitly stated to that in treatments where sanctions are implicit in the parameters of the game. To identify potential spillovers, we include two treatments that allow subjects to move across different legal environments, either sequentially from one round to the next, or simultaneously, interacting across games played in parallel. We find that legal rules do increase cooperation, and to an even greater extent than when they are implicit in the payoffs of the game. We also find evidence of spillovers across different legal environments. These results suggest that legal rules have important psychological effects on cooperative behaviour in the lab.

Contents

List of figures

List of tables

Acknowledgments

I WOULD LIKE TO EXPRESS MY SINCERE GRATITUDE to my thesis advisor Professor Emeric Henry for his continued support throughout my academic career at Sciences Po. I would also like to thank Professor Roberto Galbiati. Their input and advice has been invaluable to my study and research. My sincere thanks also goes to Sri Srikandan for his tireless work in the design and implementation of this experiment. And for lending his time and programming wizardry skills to help me overcome my LaTeX dilemmas. I would like to thank my mentor, Professor Michael Dorsch of the Central European University. Thank you Cle for keeping me company. My love and thanks to Asa for everything. And finally, I would like to thank my family for all their love and support. In particular, I express my endless gratitude to my Mother - my greatest supporter, inspiration, confidant - without whom I would never be where I am today.

When people accept breaking the law as normal, something happens to the whole society.

Orson Welles

1 Introduction

One of the fundamental challenges in economics is to understand how incentives created by institutions and formal rules (i.e. laws) affect behaviour. In particular, scholars have focused on how externally enforce rules affect cooperative behaviour.

Cooperative behaviour is a necessary condition for the development and provision of public goods. But while cooperation is socially optimal, it is individually costly; for each individual there is an incentive to free-ride. In case of private provisions of public goods, free-riding leads to inefficient underprovision. According to the prediction of the standard rational choice model, *rational, self-interested individuals* are unable to overcome collective action problems unless formal rules create incentives to act in the common good.

Economic theory commonly views law, defined as "an obligation backed by a

sanction" [Cooter, 1998], as a tool for facilitating cooperation. Traditional economic analysis has typically considered legal sanctions as market prices, to which actors adjust behaviour to equate marginal benefits to marginal costs. According to this view, legal sanctions achieve cooperation by changing incentives so that individuals are deterred from defecting. In addition to the deterrent effect of law, economists and legal scholars have emphasized the expressive function of law. The expressive theory of law states that, by expressing social values, law can create a focal point around which individuals coordinate. Moreover, law can induce individuals to change their preferences, thereby attaching a psychological penalty to defecting.

This paper studies an additional effect of law beyond its deterrent and expressive functions, which has received little attention in the current literature; namely, the spillover effect of formal institutions.

There are great variations, both geographically and temporally, and across areas of application, in the strength of legal rules. For instance, penalties for littering (a classic public goods problem) are strong in some countries while mild (i.e. non-deterrent) in others. Within a country, some aspects of contractual relationships may be enforced by legal rules while others are not. In some countries legal systems may be pluralistic: different legal systems may coexist within the same geographic area, or different laws may govern different groups of the population. In present-day Jerusalem, for example, two sharï'a courts, one Israeli and one Jordanian, operate in the city and serve the same population [Shahar, 2013].

Within a country behaviours historically regulated by customary rules may at some point be regulated by new sets of formal rules. This process of codifying informal rules occurred in Cambodia during the period of colonial rule. Before the establishment of the French protectorate in 1863, social rapport was governed exclusively by customary rules that developed over time through traditional norms. Cambodia's legacy of legal transplant is an example of dramatic change to the legal landscape associated with regime change. However, more subtle changes to legal rules resulting from government change can have important implications for compliance. For example, newly elected parties may be more or less strict on cracking

down against tax evasion - similar to the sequence of approaches to tax evasion in Italy from lenient - Berlusconi- to tough - Prodi and Monti.

These examples raise the question of whether spillovers occur from areas or tasks covered by formal mechanisms to those that are not. While theories on the extra-deterrent functions of law have been used to explain cooperation in a variety of environments, analyses tend to assume that legal rules are static. This study attempts to shed light on the dynamic effects of formal institutions, resulting from a changing legal environment across time, or from the impact of concurrently conflicting law behavioural effects of fluctuating. The question of institutional spillover effects, which has not been addressed in the literature, is central to the present study.

To address this question, we implement a series of laboratory experimentsThe current was conducted as a pilot to an ongoing research project by Professors Emeric Henry and Roberto Galbiati on the subject of institutions in infinitely repeated games.. Specifically, we examine experimental play in infinitely repeated prisoner's dilemma games in which different legal rules are applied to non-cooperative action. In the absence of sufficient observational data to overcome standard identification problems, lab experiments have become increasingly popular amongst researchers. Lab experiments have the additional advantage of providing a fully controlled environment where causal factors for behaviour can be identified. This allows us to make hypotheses about the role of legal rules in inducing specific behaviour, and to then compare them to observed the controlled variation in behaviour when specific legal rules are introduce.

In these experiments participants interact in a series of indefinitely repeated prisoner's dilemma where the "legal rules" affect the parameters of the interaction. In the repeated prisoner's dilemma participants make a clear choice between a cooperative and a non-cooperative option. In this environment we will study the following questions:

1. *What is the effect of legal rules (legal punishments for choosing the non-cooperative option) on levels of cooperation? More specifically can apparently mild sanctions play a role by changing the equilibrium play in a repeated game setting?*

2. *Do legal punishments have an additional effect other than just changing payoffs (extra deterrent effect)?*

3. *When areas are covered by legal rules while others are not, are there spillovers from one to the other? In what direction?*

The remainder of this chapter presents an overview of the experimental results; Chapter 2 reviews the current literature on the theory of institutions and experimental analysis of infinitely repeated prisoner's dilemma games; Chapter 3 discusses the prisoner's dilemma in the context of the present study; Chapter 4 presents the experimental design and procedures; Chapter 5 provides theoretical predictions to our three empirical questions from both the standard economic and recent theoretical perspectives; Chapter 6 presents the experimental results and discusses their implications; Chapter 7 concludes and suggests a future research agenda.

1.1 Overview of the results

This paper experimentally analyzes the effects of mild sanctions in a laboratory setting. We present evidence from six experimental sessions of five different treatments of repeated prisoner's dilemma (PD) games. In each stage-game, players choose to either "Cooperate" (choose C) or "Defect" (choose D)[1]. Given the payoff specifications and probability of continuation ($\delta = {}^{3}/_{4}$) in games where legal rules are absent, defection is the only ration strategy.

To test the effect of legal rules on levels of cooperation, we compare treatments in which cooperation is not supported in equilibrium with treatments where sanctions are imposed on defection. Though the parameters of the game are constant, sanctions implicitly alter the material incentives of the game so that cooperation is both Pareto efficient and risk dominant (as defined later). In all treatments where legal rules are present, we find that cooperation is indeed higher than in treatments where legal rules are absent and cooperation is not an equilibrium action. If we

[1] Neutral language was used in the experiment to denote the set available actions: X and Y, respectively.

consider cooperation rates across all rounds and all matches, subjects cooperate nearly three times more in penalty games than in those without penalties.

To test whether sanctions affect cooperation only by changing the relative payoffs, or whether they have additional psychological effect on subjects' behaviour, we study cooperation in treatments where the changes to payoffs are internalized in the parameters of the game. The equivalent payoff treatment without formal legal rules serves as a control and benchmark against which the effects of formal sanctions can be assessed. This allows us to disentangle framing and deterrent effects of formal rules on cooperative behaviour. We find that when legal rules are stated explicitly, rather than implicit in the payoff function of the game, subjects tend to achieve higher levels of cooperation. Across all matches, players cooperate 27.61% of the time under treatments in which penalties are explicit and present in every match, against 17.31% in treatments where the increase in the marginal cost of defection is implicit in the game. This evidence suggests that legal rules do have extra-deterrent effects on cooperative behaviour in the lab.

We study potential spillover effects of formal rules by examining levels of cooperation in treatments where subjects move sequentially between games with and without legal rules. To separate potential spillovers from other processes of learning with experience, we study behaviour in a state of the art treatment[2], in which games are played in parallel and players interact simultaneously in games with and without penalties. This treatment allows for potential spillovers to occur simultaneously across different legal environments. While additional research is needed to more thoroughly analyze the question of spillovers, the early evidence from this study suggests that they do in fact occur. Together, these treatments suggest that the experiences in different legal environments, occurring either sequentially or simultaneously, spillover onto subsequent interactions.

The experimental results from this study are relevant for the theory of institutions and for the experimental analysis of infinitely repeated prisoner's dilemma.

[2]The designed and implementation of this innovative treatment was achieved with the help of Sri Srikandan, IT manager and programming wizard of the Department of Economics at l'École polytechnique.

To the best of our knowledge, this paper presents some of the first evidence in the literature on *i*) disentangling framing effects of formal rules in an infinitely repeated setting, and *ii*) spillover effects of enforcement on cooperation when people change institutional settings.

Most conduct is guided by norms rather than by laws. Norms are voluntary and are effective because they are enforced by peer pressure.

Paul Collier

2

Related Works

Over the last few decades, a burgeoning experimental literature has developed that has sought to address various issues related to the effects of law and formal institutions (e.g. sanctions; subsidies) on behaviour in a laboratory setting. The experimental results presented in this paper contributes to the current literature. In particular this study relates to the following strands of the literature: the dynamics of behavioural spillovers; empirical legal and economics studies on laws and norms; the effects of institutions on the provision of public goods; infinitely repeated prisoner's dilemma.

2.1 The dynamics of behavioural spillovers

Peysakhovich and Rand [2013] suggest that the variability in cooperative norms is driven in part by spillovers from environments in which individuals predominantly interact. Based on a series of economic experiments in the lab, the authors suggest that individuals who interact in environments where cooperation is advantageous tend to develop cooperative habits. Peysakhovich and Rand [2013] show that following play in economic games where cooperation is supported in equilibrium, subjects tend to be more cooperative in one-shot games (where opportunistic behaviour is advantageous), more prosocial, more likely to punish selfishness ("norm-enforcement"), and more trusting.

Peysakhovich and Rand [2013] propose a Social Heuristics Hypothesis (SHH) to explain the emergence of cooperative norms. The SHH contends that "individuals develop heuristics[1] for social interactions" that they "carry with them to other decision settings" [Peysakhovich and Rand, 2013: 2].

The SHH predicts that social preferences are shaped by the environments in which interactions take place. When strategies (i.e. a particular set of behaviours and preferences) in our daily experiences are successful, they become internalized as norms. Spillovers from environments where cooperation is supported (by effective formal or informal institutions) are expected to make subjects more inclined towards prosocial behaviour. As Bohnet et al. [2001] demonstrate, different legal institutions may be more or less conducive to the economic success of cooperative behaviour and thus to the emergence of cooperative norms.

Peysakhovich and Rand [2013] consider spillovers from different types of interactions, in particular, from behaviour in a prisoner's dilemma to decisions in one-shot games. Their experiment only allows for uni-directional spillovers from

[1] Gigerenzer and Gaissmaier [2011] define heuristics as "efficient cognitive processes, conscious or unconscious, that ignore part of the information." To quote Gigerenzer and Gaissmaier [2011], Heuristic decisions may not conform with "'rational' decisions as defined by logic or statistical models. However, for many decisions, the assumptions of rational models are not met, and it is an empirical rather than an a priori issue how well cognitive heuristics function in an uncertain world" [Gigerenzer and Gaissmaier, 2011: 451].

norms forming either in cooperative or non-cooperative environments to environments where cooperation is not optimal. While their approach offers an effective way of isolating behaviour, it does not allow for dynamic and simultaneously occurring spillovers. We instead consider the different environments in which they might occur, and hold the type of interaction constant (i.e. play in repeated prisoner's dilemma). This allows us to isolate the introduction of legal rules as our variable of interest. Our approach is also unique with respect to the experimental design, which includes two treatments dedicated to the question of spillovers. As discussed briefly in Chapter 1, we study behaviour in treatment where the presences and sequence of legal rules is determined randomly. In a final treatment, players are organized into trios with one participant playing two game simultaneously in parallel with the other two players: one involving legal sanctions and one without. By comparing treatments where players are both sequentially and simultaneously exposed to different legal environments, we can isolate any potential spillover effect. To the best of our knowledge, no other studies have executed simultaneous prisoner's dilemma game played in parallel to effectively control for spillovers from sequential play.

In the context of compliance with legal rules, Bruttel and Friehe [2010] presents experimental evidence of spillovers in tax compliance behaviour. For given values of the audit probability and severity of the fine for tax evasion, the authors asses the level of income declaration of subjects who face a change in the enforcement of tax laws. Bruttel and Friehe [2010] demonstrate that past tax enforcement regimes continue to have an impact on current income declarations. In particular, they find that subjects who initially experience weak enforcement tend to declare less income when faced with strict enforcement than subjects who have only experienced strict enforcement; subjects who initially experience strict enforcement tend to declare more income when faced with weak enforcement than subjects who have only experienced weak enforcement [Bruttel and Friehe, 2010].

In a fascinating natural experiment based on data on parking violations among international diplomats living in New York City, Fisman and Miguel [2006] examine the importance of legal enforcement versus cultural norms in mitigating

corruption. The researchers examine differences in behaviour of government employees from different countries and domestic legal systems, all of whom can act with impunity in illegally parking their cars while in New York City. Prior to 2002 diplomatic immunity allowed consular personnel to avoid paying parking fines, effectively creating a zero enforcement environment with non-deterrent local parking laws.

The propensity of diplomats across difference societies "to break rules for private gain when enforcement is not a consideration," [Fisman and Miguel 2006: 1] provides a natural measure of the spillover effect of different legal rules on behaviour.

Standard economic models predict that in the absence of legal enforcement, all diplomats should have a high propensity to violate local parking laws. However, Fisman and Miguel [2006] find that diplomats from countries where laws against corruption are strictly enforced have significantly fewer parking violations (and tend to pay them) than diplomats from high corruption countries. This observation suggest that "they bring the social norms ... of their home country with them" Fisman and Miguel [2006]. In other words, there are spillovers from the legal environment of their countries of origin. Moreover, spillovers from countries where cooperative behaviour is advantageous (i.e. supported in equilibrium) are found to be particularly strong.

Fisman and Miguel [2006] also examine the evolution of diplomats' behaviour during their tenure in New York City. They find evidence of learning, with the frequency of unpaid violations increasing "rapidly and statistically significantly with tenure, " converging towards the "zero enforcement" equilibrium [Fisman and Miguel 2006: 11]. The researchers find a positive interaction effect of home-country corruption and tenure, "implying there is a divergence in the rate of violations over time, with diplomats from high corruption countries showing increasingly more violations relative to diplomats from low corruption countries over time perhaps as they learn about the reality of diplomatic immunity" [Fisman and Miguel 2006: 11]. This process of equilibrium-convergence of behaviour is approached experimentally by Dal Bó and Fréchette [2011].

The study also highlights the effect of a sudden change in legal rules, in this case, the implementation of the Clinton-Schumer Amendment in October 2002. With the enactment of the law, "proposed by the Bloomberg administration to deal with the diplomat parking problem," Diplomats impunity was withdrawn [Fisman and Miguel 2006: 5].

During the post-enforcement period, overall levels of parking violations fell dramatically. The authors interpret this effect as suggesting that "increased enforcement can sharply reduce corruption" [Fisman and Miguel 2006: 5]. However, this interpretation says little about the underlying behavioural effects of the law.

2.2 Laws and norms

The standard economic theory of compliance with law emphasizes the direct deterrent effects of legal sanctions, which is based on the assumption that individuals behave rationally to maximize their utility. According to this view, people weight the costs and benefits of breaking the law and incurring legal sanctions.

In the context of the Clinton-Schumer amendment described by Fisman and Miguel [2006], the deterrence effect of the law induces compliance by altering the gains to committing parking violation. When enforced, compliance local parking laws is economically advantageous, and it is no longer rational to violate them. However, deterrence theory fails to explain the behaviour the diplomats who chose to comply with the non-deterrent laws during the zero enforcement period, when it was not economically rational to do so.

Recent thought in law and economics has attempted to explain these behaviours, emphasizing the importance of the indirect effects of law on social norms. This line of reasoning implies that their their impact on social norms, laws may influence behaviour even in the absence of deterrent sanctions.

Social norms are commonly defined as "informal social regularities that individuals feel obliged to follow because of an internalized sense of duty, because of a fear of external non-legal sanctions, or both" [McAdams, 1997: 340]. In other words, social norms are a set of internalized constraints. Scholars have suggested that by

expressing social values, laws may induce rational individuals to internalize social norms. Tyran and Feld [2006] note that laws may encourage the emergence of cooperative norms if the law is perceived as a public expression of what one ought to do. The law may not only represent a price to pay for acting a certain way, but may also signal that such behaviour is unacceptable. The idea is that laws express the reigning norms in a society, and can discipline people by showing them what the majority of people deem to be 'appropriate'. That is, "by prohibiting an action or declaring it as a duty, law can weaken the norm it condemns and strengthen the one it embodies" [Funk, 2007: 135]. This literature explains the effects of mild sanctions that, according to rational choice theory, should not affect behaviour in non-repeated environments [Kahan, 1997].

The literature discusses several channels through which this expressive function of the law can take effect. On the one hand, laws may shape preferences and "elicit intrinsic motivation by framing an act as wrong" [Bohnet and Cooter, 2003]. Alternatively, law may coordinate the behaviour of different people by changing their beliefs about what others will do. In particular laws may induce cooperation through the emergence of the norm of *conditional cooperation*, which prescribes that one ought to cooperate if others also cooperate. Ostrom [2000] argues that conditional cooperation may be the most important social norm in motivating cooperative behaviour. As a direct public expression of appropriate behaviour, laws may indirectly induce conditional cooperation. In other instances laws may be counterproductive and actually "crowd out" cooperation [Bohnet et al., 2001, Gneezy and Rustichini, 2000].

Consider the classic public goods problem of littering. Given that the expected value of fines for littering are typically quite low (i.e. that anti-littering ordinances are an example of a mild law), it is surprising from an economic perspective that not all people litter. However, some people are intrinsically motivated not to, even in the absence of anti-littering ordinances. Others may litter in the absence of specific laws, but refrain from doing so out of intrinsic respect of the law if they are present. Conditionally cooperative individuals may choose not to litter if it is perceived to

be socially inappropriate [2]. In this respect, even non-deterrent anti-littering ordinances may contribute to cleaner streets through their effects on the emergence of cooperative norms.

Nobel Laureate Thomas Schelling (1960) first observed that, in many real would situations with multiple equilibria, behaviour is not uniquely determined by material payoffs. In situations "requiring coordination, anything that makes salient" or "draws the mutual attention of the individuals who need to coordinate," "tends to produce self-fulfilling expectations that this equilibrium will occur" [McAdams, 2009, p. 26]." In this sense, the expressive function of law may serve as an equilibrium selection principle by coordinating expectations [McAdams, 2000, McAdams and Nadler, 2005]. The tendency of subjects to gravitate toward the focal outcome in situations with multiple equilibria (e.g. infinitely repeated prisoner's dilemma), means that lawmakers can use legal rules to increase the likelihood of compliance by making focal the form of behaviour the law demands.

Several theorists have presented evidence of the law working in this manner. Bohnet and Cooter [2003] show how very mild (non binding) penalties for choosing the inefficient strategy in a coordination game induce more people to choose the efficient strategy. Bohnet and Cooter [2003] experimentally investigate the framing and coordination effects of laws on behaviour in prisoner's dilemma, "crowding" and coordination games. They compare games with unique Nash equilibria, where a sanction can only affect behaviour by changing preferences, and games with multiple equilibria, where a sanction can help to select among them. Bohnet and Cooter [2003] find that a penalty does serve as an equilibrium selection in coordination games but has no effect on subjects' preferences in prisoner's dilemma or crowding games.

Kahan [1997] expands upon the theory of extra-deterrent law by suggesting that individuals' propensity to commit crimes is determined in part by the perception that criminal activity is widespread. The author suggests that law not only deters criminal activity by affecting the "price" of crime - the severity of punishment dis-

[2] Evidence from controlled field studies reveals that people tend to litter significantly less in clean environments [Krauss et al., 1978].

counted by the probability that it will be imposed - as the standard economic conception assumes, but that it also affects individuals' perceptions of the conduct and values of other individuals. Consequently, some policies that appear to be effective means of raising the price of crime might in fact undermine deterrence by magnifying the perception that crime is rampant.

Funk [2007] provides the first empirical evidence of the expressive function of law. The author studies the effect of the abolishment of mandatory voting laws in Switzerland (previously enforced with a fine amounting to no more than one dollar) on voter-turnout. She finds that the legal abolishment of the voting duty significantly decreased average turnout, even though the fines for not voting had only been symbolic. The author suggests that the effect of abolishing the fine changed the public perception of the moral incentive to vote. The legal declaration of voting as a duty apparently caused high voter-turnout, which Funk [2007] interprets as evidence of the expressive effect of law.

Tyran and Feld [2006] run an interesting experiment that compares the compliance-inducing effects of endogenously and exogenously introduced 'mild' or 'non-deterrent' sanctions in a public good game[3]. In the endogenous treatment, the subjects vote on whether to introduce the sanction. The interpretation of this result is that endogenous sanctions signal that there are many people who want to cooperate. The authors also test the hypothesis that exogenous non-deterrent law activates norms by expressing what one ought to do.

The authors show that endogenous sanctions are more effective in raising contributions than exogenously implemented sanctions. Voting to enact the law induces the expectation of cooperation. These results experimentally demonstrate the social influence conception put forward by Kahan [1997] which suggests that people tend to comply with the law if they expect others to do the same.

Bohnet et al. [2001] study how behaviour evolves in different institutional setting with respect to contract performance. Bohnet et al. [2001] analyze play in re-

[3]By "non-deterrent", the authors are referring to sanctions that do not fully deter free-riding, whereas when the law is "severe", a rational and self-interested agent will fully contribute to the public good.

peated contract games with varied probabilities of contract enforcement, in which players have an opportunity to maximize their payoffs by entering into contract. The first mover in the game decides whether or not to offer a contract. The second mover then decides if she will perform or breach. Without knowing if her partner will breach the contract, player 1's decisions is thus a matter of trust [Bohnet et al., 2001].

The interaction is played out under three different legal regimes with varying degrees of contract enforcement. Standard economic theory of law predicts "the higher the expected cost of the breach, the more likely is the second mover to perform," [Bohnet et al, 2001: 131]. However, [Bohnet et al., 2001] find a nonmonotonic relationship between behaviour and the probability of enforcement (Bohnet et al, 2001: 131). In other words, performance rates on contracts are high when enforcement is either sufficiently strong or sufficiently weak. When the probability of enforcement is only moderate, second movers tend to breach on average.

Bohnet et al. [2001] interpret these results as reflecting the evolution of individuals' preferences (i.e. trustworthy individuals demonstrate a preference for honesty and suffer psychological costs from breaching), which are more resilient when they are economically advantageous. Legal rules may alter the parameters of an interaction in such a way that make either cooperative of opportunistic behaviour more economically successful. Consequently, legal rules can crowd-out as well as crowd-in preferences. Bohnet et al. [2001] find that weak contract enforcement crowds-in preferences for honesty and trustworthiness, while intermediate enforcement crowds-out reciprocity-driven behaviour. When the probability of enforcement is high, the second mover is deterred from breaching regardless of prevailing preferences. In this case institutional trust is said to perfectly substitute personal trust [Frey and Jegen, 2001].

When legal institutions are inefficient, a sufficiently informative signal of a player's trustworthiness can serve as a substitute for strong legal enforcement. On average, honest individuals will be more economically successful as they receive more contracts than dishonest individuals. Trustworthiness is therefore crowded in.

Intermediate levels of enforcement have the reverse effect of crowding-out trust-

worthiness, in favor of opportunistic behaviour. The expected payoff is always higher from entering into a contract. Therefore, players always prefers to offer a contract, regardless of perceived trustworthiness of their opponent. Such indiscriminate behaviour means that dishonest players can maximize their expected payoffs by taking advantage of profitable opportunities to breach. As a result, honesty is less economically successful and is crowded out.

The canonical example of institutional crowding out is provided by Gneezy and Rustichini [2000], who conduct a field study on parents' responses to the introduction of a fine for arriving late to pick up their child from day-care. The results of the experiment offer insight into the effects of monetary incentives on behaviour. While the deterrence hypothesis predicts that"the introduction of a penalty that leaves everything else unchanged will reduce the occurrence of the behaviour subject to the fine," Gneezy and Rustichini [2000] observe the opposite effect[Gneezy and Rustichini, 2000: 1]. With the introduction of the fine, the number of late parents significantly increased. Furthermore, after the fine was subsequently removed, the frequency of late arrivals did not immediately return to pre-fine levels, but remained stable at the higher rate for the next four weeks.

These findings are interpreted as suggesting that the fine was perceived as merely a price for being late [Fehr and Gächter, 2002]. The introduction of the fine "transformed the act of being late from a rule violation to a market transaction" [Fehr and Falk, 2002: 26]. In the pre-fine period, the rules regarding late arrivals were part of an implicit (albeit incomplete) contract made with the staff of the day-care centers [Gneezy and Rustichini, 2000]. Parents faced a non-pecuniary cost of disapproval for being late. The higher frequency of late arrivals with the introduction of the fine suggests that the cost of disapproval exceeded the amount of the fine. Moreover, the fine changed the perception of the parents regarding the environment in which they were operating and its associated appropriate behaviour [Gneezy and Rustichini, 2000]. The results of the experiment shed light on the interplay between social norms and monetary incentives.

As a final point, these dynamics between social norms and formal rules can also be explained from a evolutionary biological perspective. According to evolution-

ary theory, "the human brain appears to have evolved a domain-specific, human-reasoning architecture, rather than developing general analytical skills that are applied to a variety of specific problems" [Clark and Karmiloff-Smith, 1994]. As such "humans use a different approach to reasoning about deonic relationships - what is forbidden, obligated, or permitted - (where they tend to check for violations, or cheaters) as contrasted to reasoning about empirical relationships - what is true and false- (where they tend to use a confirmation strategy)" [Ostrom, 2000]. In other words, modern humans have inherited a propensity to learn social norms.

2.3 The effects of institutions on the provision of public goods

This study is particularly relaxant to the literature focusing on exogenously imposed institutions. In this strand of literature, Falkinger et al. [2000] examine a mechanism for public good provision in which rewards and sanctions are imposed to players who contribute more and less, respectively, than the average to the public good. The authors show experimentally that the formal mechanism is an effective tool to achieve efficient contribution levels. Andreoni [1993] presents an experimental test of the proposition that government contributions to public goods, funded by lump-sum taxation, will completely crowd-out voluntary contributions. While both these papers show that it is possible to design exogenous institutions that can efficiently sustain cooperation, a related large strand of literature focuses on the crowding-out effect of incentives.

Fehr and Gächter [2002] and Bowles and Polaní-Reyes [2012] provide an extensive survey of the literature on the psychological forces the influence the effectiveness of economic incentives. They emphasized that tendency of economic theory to focus on the desire to avoid risk and to achieve income in the modeling of incentives, neglects the importance of powerful non-pecuniary motive that shape human behaviour. The authors discuss the effects of three such motives - the desire to reciprocate, the desire to gain social approval, and the intrinsic enjoyment arising from working on interesting tasks. These motives interact with pecuniary

incentives in ways that may undermine or complement their intended effect. In particular, these psychological forces may cause monetary incentives to backfire and reduce compliance of agents with formal rules. Conversely, such motives may generate powerful incentives themselves that complement material incentives to behave in a way that law demands.

This paper complements this literature by introducing the question of spillovers of legal rules.

2.4 Experimental analysis of infinitely repeated prisoner's dilemma

Finally this study will contribute to the recent experimental literature on cooperation in infinitely repeated games[4] (early examples include: Roth and Murnighan [1978], Palfrey and Rosenthal [1994] and more recently: Dal Bó [2005]; Andreoni [2009]; Dal Bó and Fréchette [2011]; Bigoni et al. [2012], Benoit et al. [2013] among others).

Roth and Murnighan [1978] conducted the first experimental study of equilibrium behaviour in repeated prisoner dilemma (PD) games. They introduced the method for inducing infinitely repeated games in the lab using a random continuation rule. "More precisely, after every round, the game ends with probability $1 - \delta$ and the next round formed by the same stage game continues with probability δ" [Blonski et al, 2007, p. 8]. Since its introduction, a large number of studies have followed this method of interpreting the infinitely repeated game as one with stochastic termination. Roth and Murnighan [1978] show that cooperation is greater when it can be supported in equilibrium and increases with the probability of continuation.

However, Dal Bó [2005] identifies several methodological problems with the experimental design of Roth and Murnighan [1978]: subjects play against the experimenter rather than each other; the experimenter restricts play to two strategies

[4] A more detailed discussion of the prisoner's dilemma itself is provided in the following chapter.

(Tit-for-Tat or Grim Trigger), unnecessarily influencing subjects' behaviour based on previous play; subjects are not paid proportionately (10$ for the "best player").

Murnighan and Roth [1983] perform a similar test of the effect of the expectation of continued play using twelve variations of the PD game. They find that for particular parameters of the game, as the probability of continuation increases, so too does cooperation. They conclude that the expectation of continuation, interacted with the given payoffs, is an important determinant of cooperative behaviour.

Dal Bó [2005] considers whether the responsiveness of cooperation to an increase in the continuation probability observed by Roth and Murnighan [1978] is simply a reflection of the increase in the expected number of repetitions of the stage game before a relationship is terminated, or whether behaviour in indefinite games is fundamentally different from behaviour in finite repetitions of a PD stage game. To test this, Dal Bó [2005] examine behaviour in infinitely repeated games with various probabilities of continuation, controlling for subjects' behaviour in finitely repeated games of the same expected length. He finds that the percentage of cooperative play in finitely repeated games of a given length is significantly lower than in its indefinitely repeated counterpart. This result leads the authors to conclude that "the shadow of the future" does indeed impact cooperation as theory predicts [Dal Bó, 2005].

Dal Bó and Fréchette [2011] examine how cooperation in infinitely repeated Prisoners' Dilemma games evolve with experience. This is motivated by the observation of Dal Bó [2005] that while cooperation as an equilibrium action is a necessary condition of the level of cooperation to increase with experience, it does not appear to be a sufficient condition. This raises the question of the underlying determinants of cooperation and the criteria for equilibrium selection in infinitely repeated games.

They simulate infinitely repeated games using a random continuation rule. Their main treatment variables of interest are the probability of continuation and the payoff from cooperation. The authors apply the concept of strategic risk to their analysis, following Blonski and Spagnolo [2001]. They find that as subjects learn,

they may reassess the value of 'investing' in cooperation and thus change their own behaviour accordingly. The perceived gains from cooperation are found to be determined in part by the fundamental parameters of the game, the behaviour of players' previous opponent, and the realized length of the repeated game.

Blonski et al. [2007] demonstrate that risk dominance is an important determinant of whether cooperation emerges in infinitely repeated games. They find that the condition of risk dominance is strictly stronger than the presence of cooperative equilibria. Dal Bó and Fréchette [2011] qualify this result, noting that while cooperation as an equilibrium and risk dominant action are found to be necessary conditions for cooperation to increase with experience, they are still not sufficient.

However, with sufficiently high payoffs to cooperation and probability of continuation, subjects achieve very high levels of cooperation with experience.

Duffy and Ochs [2009] study the impact of matching protocols (fixed or random) on the emergence of a cooperative norm to test Kandori's [1992] theory that "a social norm of cooperation can be sustained as a non-cooperative equilibrium even when individual pairings are both random and anonymous" [Duffy et al., 2009: 786]. Duffy and Ochs [2009] examine play in a series of two-stage repeated prisoners' dilemmas in which players switch from games with fixed matching, where cooperation is in equilibrium, to games with random matching, where it is not. Within the random matching treatment players are exposed to varying information signals regarding the past behaviour of their opponents. Though the switch to the second phase if found to induce a large reduction in cooperation as theory predicts, there is some indication of a spillover effect. Subjects cooperate significantly more in the first round of the second stage where cooperation is not supported in equilibrium than in subsequent periods as they quickly adapt to the new environment. Contrary to Kandori [1992], Duffy and Ochs [2009] find that no cooperative norm emerges in the random matching treatment, but that cooperative norms do emerge with fixed pairings as subjects gain experience. Informal mechanisms prove insufficient to enforce cooperation amongst anonymous, randomly matched subjects. This leads the authors to conclude that formal institutions that make individuals' reputations public are necessary to sustain cooperative

relationships.

While this literature has generally focused on understanding both the dynamics and conditions that favour cooperation in infinitely repeated games, Benoit et al. [2013] focus on the comparison of investment choices under different institutional regimes. The present study examines the effect of imposing sanctions in this type of environment and how they might generate spillovers.

all science, all human thought, is a form of play.

William Poundstone

3
The Prisoner's Dilemma

THE CLASSIC PRISONERS' DILEMMA has come to be viewed as the canonical representation of cooperative problems where tension exists between what is individually optimal and what is socially optimal.

McAdams [2009] vividly illustrates the iconic narrative from which the game gets its name:

> *A prosecutor suspects two prisoners of a felony, but can currently prove their involvement only in a misdemeanor. The prosecutor offers each prisoner the same inducement to confess to the felony, (summarized in the payoff matrix below): 'If you are the only one to confess, I will reward you by dropping all charges,' (payoff of c). 'If you are the only one not to confess, I will use your confederate's testimony to convict you of the felony and obtain for you the maximum [five] years in prison (a); if neither of*

you confesses, you each get [one] year for the misdemeanor (c); if both confess, I will convict you both of the felony but give you an intermediate sentence of [three] years (d).'

In this context, the strategy of not confessing is to *cooperate* and the strategy of confessing is to *defect*. Altruism can alter the outcome of the game, but the standard assumption is that the prisoners are self-interested and rational, caring only about their own punishment.

The symmetric prisoners' dilemma stage game is represented by the payoff matrix,

		Player 2: C	Player 2: D
Player 1	C	c, c	b, a
	D	b, a	d, d

characterized by payoff parameters a, b, c, d such that $b > c > d > a$ and $2c > b + a$[1]. The first entry in the cell represents Player 1's payoff, while the second entry represents the payoff of Player 2. Players simultaneously choose whether to cooperate (C) or deviate (D)- they do not observe the choice of their opponent before choosing. The choices made by each player determine their payoffs in a given round according to the matrix above.

In the one-shot Prisoners' Dilemma (PD), the game has a unique *equilibrium*, an outcome where each player chooses her best response to the strategy of her opponent. More precisely, an equilibrium refers to a *Nash equilibrium*, the central

[1]The first parameter defines the classical form of the PD while "the second restriction excludes cases where indefinite cooperation is not efficient since patient players can improve by defecting alternatively" [Blonski and Spagnolo, 2001: 6 note 10]

solution concept in game theory, "based on the principle that the combination of strategies that players are likely to choose is one in which no player could do better by choosing a different strategy given the ones the other choose. A pair of strategies will form a Nash equilibrium is each strategy is one that cannot be improved upon given the other strategy" (i.e. neither player has any incentive to deviate) [McAdams, 2009: 4 note 9].

According to the payoff matrix above[2], if Player 2 cooperates, Player 1 is better off deviating and receiving a payoff of b than cooperating and receiving a (since by definition $b > a$). If Player 2 deviates, Player 1 is better off deviating as well (receiving d rather than a from cooperating). Therefore, since the payoffs are symmetric, both Player 1 and Player 2 have a *dominant* strategy of deviating; it is their best move regardless of what their opponent does. Thus, in the one-shot PD, the only Nash equilibrium is to play (D, D). Yet because $c > d$, each player regards this outcome as worse than the outcome (C, C). Hence, the game is termed a 'dilemma' because the theoretically inevitable outcome (D, D) is worse for both players than if they were able to coordinate to cooperate (play (C, C)).

One mechanism that can sustain mutually beneficial cooperation in the PD is repetition of the standard one-shot game. "This result plays out in the fact that each player fears that one instance of defecting will lead to a collapse of cooperation in the future. If the value of future cooperation is large and exceeds what can be gained in the short term by defecting, then the long-term individual interest of the players can automatically and tacitly keep them from defecting, without the need for any additional punishments of enforcement by third parties" [Dixit et al., 2009: 399].

3.0.1 Cooperation criterion in the prisoner's dilemma

In the simple form of the PD, two players have two discrete actions available to them, which they take simultaneously. Let $G^{\infty}(\delta)$ be the two-player infinitely repeated simultaneous-move stage game, where players discount future payoffs pro-

[2] In each cell of the matrix, the payoffs for Player 1 are on the left and the payoffs for Player 2 are on the right.

portionately at a common discount rate $\delta \varepsilon [0, 1)$. Each stage game consists of a finite action space A_i for each player and a payoff function $g_i : A \to \Re$, for $i = 1, 2$, where $A = A_1 \times A_2$ and $\Re$ is the set of real numbers. The payoff for each player is determined by the strategy choices of both players, such that $g_i(a_i{}^t, a_{-i}{}^t)$ is the stage payoff to player i when the action profile $a^t = (a_i{}^t, a_{-i}{}^t)$ is played.

The sequential nature of the relationship in repeated games means that players can adopt strategies that depend on behaviour in preceding periods. Let $\mathbf{a} = \{a^t\}_{t=0}^{\infty}$ denote the infinite sequence of action profiles. The game is assumed to begin in period $t = 0$ with the null history h^0. In period t, the history of previously realized actions is given by $h^t = \{a^0, \ldots, a^{t-1}\}$, and the set of all possible period-t histories is H^t.

All players observe the outcomes of all past periods (*i.e. perfect monitoring*). Therefore, a pure strategy s_i for player i is a strategy set $s_i{}^t$ that, for each period t, maps all possible period-t histories $h^t \varepsilon H^t$ to actions $a_i \varepsilon A_i$. That is, $s_i = \{s_i{}^t\}$, where $s_i{}^t : H^t \to A_i$.

Player i's discounted payoff for the entire repeated game is the sum,

$$u_i(\mathbf{a}) = \sum_{t=0}^{\infty} \delta^t g_i(a_i{}^t, a_{-i}{}^t) \tag{3.1}$$

where, again, $\delta \varepsilon [0, 1)$. The discount factor δ can be thought of as representing pure time preference, such that $\delta^{-r\Delta}$, where r is the rate of time preference and Δ is the length of the period. As δ increase, player are more concerned with payoffs in future rounds relative to those in the current period. The discount factor also represents the probability that the game will continue for at least another period.

A player's *continuation payoff*, the expected payoffs from period t on, is given by,

$$\sum_{t=1}^{\infty} \delta^{t-1} g_i(a_i{}^t, a_{-i}{}^t). \tag{3.2}$$

If the prisoner's dilemma is repeated for only a finite number of times, then through backwards induction we find that (*D,D*) is the unique *subgame perfect*

Nash equilibrium (SPNE). In the last period, defections is the dominant strategy, regardless of the history of the game. Anticipating that one's opponent will defect in the last period, (D,D) is also the dominant strategy in the preceding period. Continuing inductively, (D,D) is a unique SPNE for each period, regardless of history. In other words, rational players will reason that deviation is the equilibrium in the last round, and because they expect their opponent to deviate in the last round, they also expect her to deviate in the second-to-last round, and eventually through backward induction they will work their way to the decision to deviate in the present period.

However, if the PD is repeated infinitely, *trigger strategies* enable players to threaten their opponents with a *punishment* action if they deviate from an implicitly agreed upon action profile. Infinitely repeated PD games are solved by describing combination of strategies that form subgame perfect Nash equilibria and by considering possible subgames that could arise both on and off the equilibrium path. Repeated interactions enable players to punish deviation by threatening not to cooperate in the future. Such punishment results in lower payoffs in the future that may offset any short-term gains from deviation. Thus, in $G^{\infty}(\delta)$, cooperation is supportable as equilibrium behaviour if the long-run payoff from playing C,C exceeds than the payoff from a single deviation followed by indefinite defection of both players. More precisely, player i's repeated-game strategy $\bar{s}_i = (\bar{s}_i{}^0, \bar{s}_i{}^1, \ldots)$ is the sequence of history-dependent stage-game strategies such that in period t and after history h^t,

$$\bar{s}_i{}^t(h^t) = \begin{cases} C, & h^t = \left((C,C)^t\right) \\ D, & \text{otherwise.} \end{cases}$$

Such a strategy is referred to as a *grim trigger strategy*. The determinants and boundary conditions of cooperation in stochastically repeated games have been well characterized both experimentally [Dal Bó, 2005, Dal Bó and Fréchette, 2011, Fudenberg et al., 2012, Rand and Nowak, 2013] and theoretically [Fudenberg and Maskin, 1986, Mailath and Samuelson, 2006, Blonski et al., 2007]. Additional mechanisms that can be used to sustain cooperation include sanctions, reputation,

partner choice, intergroup competition, and formal institutions [Fudenberg and Maskin, 1986, Ostrom, 1990, Fehr and Gächter, 2000, Rand and Nowak, 2013].

Consider the payoff matrix above. Suppose that up to time t, D has never been played. Then the respective payoffs to cooperation and defection are:

$$\begin{aligned} \text{Play } C \quad &\Rightarrow \quad c + c\delta + c\delta^2 + \ldots = \frac{c}{1-\delta} \\ \text{Best Dev. } D \quad &\Rightarrow \quad b + d\delta + d\delta^2 + \ldots = b + \frac{\delta d}{1-\delta} \end{aligned}$$

Cooperation is therefore supportable in equilibrium if,

$$\begin{aligned} \frac{c}{1-\delta} &\geq b + \frac{\delta d}{1-\delta} \\ \delta\left(\frac{c}{1-\delta} - \frac{d}{1-\delta}\right) &\geq b - c \\ \text{long-run incentive to cooperate} &\geq \text{short-run incentive to defect} \\ \delta &\geq \frac{b-c}{b-d} \end{aligned}$$

The *long-run incentive to cooperate* $\frac{\delta(c-d)}{1-\delta}$, as defined by Blonski et al. [2007] "is the maximal difference between equilibrium continuation payoffs". Comparing this incentive to the *short-run incentive to defect* $b - c$ yields the critical value $\underline{\delta} = \frac{b-c}{b-d}$ over which cooperation can be supported in equilibrium.

Note that the lower bound $\underline{\delta}$ does not take into account the parameter a as a determinant of players' propensity to cooperate. This non-deviation condition does not take into account "that a player who plans to cooperate face the problem that the opponent may not plan to cooperate and even gains by a failed attempt to build up cooperation" [Blonski et al, 2007:. 6]. The risk of cooperation depends largely on the value to the parameter a. For instance, as a goes to $-\infty$ it becomes increasingly risky to cooperate in any given period. For this reason Blonski et al. [2007] present a new critical value for the discount factor δ^* based on the notion of *risk*

dominance, introduced by Harsanyi and Selten [1988]. "In 2×2 symmetric coordination games an equilibrium is risk dominant if its equilibrium strategy is a best response to a mixture that assigns probability of one-half to each strategy by the other player" [Dal Bó and Fréchette, 2011]. According to Harsanyi and Selten [1988] risk dominance in games with two strict pure strategy equilibria can be evaluated by comparing the *Nash-products* of the equilibria[3] Given the difficulties (for mathematical reasons) of applying Harsanyi and Selten's definition of risk dominance to all possible strategies in infinitely repeated games, Blonski and Spagnolo [2001] focus on a simplified version of the game consisting of only two strategies: the *grim trigger* strategy (G) and the *always defect* strategy (AD). They say that cooperation is risk dominant if playing G is the best response to the other player's choosing G or AD with equal probabilities.

By comparing all cooperation equilibria with the defection equilibria, Blonski et al. [2007] derive a measure that captures players' strategic risk[4]. Consider the *short run disincentive to cooperate* defined by $-(a-d) = d-a$, which reflects the payoff loss from a failed attempt to cooperate rather than safely defecting. Adding this to the short run gains from defection yields the *total short-run incentive to defect* $(b-c)+(d-a)$. Comparing the total short run incentives with the long run incentive to cooperate yields,

$$
\begin{aligned}
\text{long-run incentive to cooperate} &\geq \text{short-run incentive to defect} \\
\delta\left(\frac{c-d}{1-\delta}\right) &\geq b-c+d-a \\
\delta &\geq \frac{b-c+d-a}{b-a}
\end{aligned}
$$

[3]"The Nash-product measures the product of both players disincentives not to behave according to the equilibrium under consideration" [Blonski and Spagnolo, 2001]. In our example, where $b > c > d > a$, (D,D) risk dominates (C,C) if $(d-a)(d-a) > (c-b)(c-b)$.

[4]for a detailed derivation of the critical value δ^* *see* Blonski et al. [2007].

and thereby provides a lower bound $\delta^* = \frac{b-a-(c-d)}{b-a}$ on discount factors below which it is risky for players to cooperate. Note that $\delta^* \to \underline{\delta}$ for $a \to d$ and $\delta^* \to 1$ for $a \to -\infty$. Since $b > c > d > a$ implies $\delta^* > \underline{\delta}$, if $\delta^* > \delta > \underline{\delta}$, cooperation is supportable in equilibrium and payoff-dominates defection but players nevertheless are unwilling to cooperate. All cooperation equilibria of the repeated PD-game $G^{\infty}(\delta)$ are risk dominant if and only if $\delta > \delta^*$. Dal Bó and Fréchette [2011] find that both payoff dominance and risk dominance are necessary conditions for levels of cooperation to increase with experience[5].

If the parameters of the game are such that repetition does not sustain cooperation in equilibrium (i.e. the incentive to deviate to too large), external penalties can be inflicted by a third party on players when they deviate. "When the payoffs have been altered to incorporate the cost of the penalty, players may find that the dilemma has been resolved [Dixit, 2009: 409].

[5]However, neither conditions were found to be sufficient.

The proper study of mankind is the science of design.

Herbert Simon

4 Experimental Design

We reproduce infinitely repeated games in the lab using a standard procedure involving a random continuation rule[1]. These games were implemented in a series of experimental sessions (6 in total) in the experimental economics lab of l'École polytechnique.

4.1 General procedures

While the infinitely repeated prisoner's dilemma is a theoretical model that is impossible to implement literally in an experimental setting, it can be proxied by games with stochastic termination rules. "A game with an uncertain end is equivalent to the infinitely repeated game with a discount factor equal to the probability

[1] *See* Dal Bó and Fréchette [2011] for a recent example.

that the game will continue in any period" [Baird et al., 1994: 168]. At the end of each interaction (or "round"), the computer randomly determines whether or not another round is played in the in the current repeated game (or "match"). The probability of continuation is fixed at 0.75 for all treatments and is independent of any choices players make during the game. Participants therefore play a series of games of random length (with expected length of 4 rounds)[2].

At the beginning of each match, players are randomly assigned to a partner. All interactions between players take place through computer terminals. Therefore, players do not know with whom they are paired. This allows us to control for the possibility of reputations or informal sanctions influencing behaviour.

In T1, and the games played without penalty in T4 and T5, defection is a symmetrically optimal strategy. In T2 and T3, and the penalty games T4 and T5, cooperation is a symmetrically optimal strategy.

In the treatments played with penalties, players who choose to defect incur sanctions equal to $aF_i = 10$ with certain probability (i.e. $a = 1$). Instructions given to participants use relatively neutral wording as a common practice in experimental economics. The cooperative actions is labelled *X* while defection is labelled *Y* and the sanctions is referred to as a "penalty".

The experiment terminates once the round being played at the 15th minute ends.

4.2 Description of experimental treatments

We implement a laboratory experiment based on the following 5 treatment that will allow us to answer the questions posed in Chapter 1:

***Base game* (T1)** In this treatment we implement a standard infinitely repeated prisoner dilemma (with two choices, *Cooperate* or *Defect*). In the repeated game, cooperation is not part of a sub-game perfect equilibrium. *This represents the no legal rule environment.*

[2]The expected length of a game is given by $\frac{\delta}{1-\delta}$.

***Penalty* (T2)** Same as in T1 but with a 'fine' imposed for playing defection instead of cooperation. *This represents the legal rule environment.*

***Payoff equivalent* (T3)** This treatment has the same payoff structure as T2 but the 'fine' instead of being introduced explicitly is integrated directly in the payoffs.

***Random - Duo* (T4)** at the beginning of each game the computer randomly determines whether subjects play T1 or T2.

***Trio* (T5)** participants play two repeated games in parallel with different partners. One played without penalty, as in T1, while the other is played with penalty, as in T2. In practice, the participants have split screens corresponding to each separate game.

Comparing T2 and T1 provides information on the effect of mild sanction in this environment. Comparing T3 and T2 disentangles framing effects of sanctions. T4 and T5 allow for an analysis of the question of spillovers. These treatments mimic in the lab two different situations where institutional spillovers might be relevant. First, T4 mimics a situation of rules changing in the same area, or move sequentially from areas covered by different legal rules. T5 represents a situation where an individual interacts simultaneously in two physical areas or over two different tasks covered by different rules. We can think of this as an individual conducting business internationally, in countries where different contract laws apply to the same activity. Or as a situation in which conditions of legal pluralism result in concurrent conflicting legal rules.

In the fourth treatment, players interact in a sequence of games that can be one of two different types: games with legal rules (corresponding to T2) and games without (corresponding to T1). At the start of each round, a random draw determines if T1 or T2 is played. Thus, playing T1 or T2 in any given round is equiprobable.

At the end of each round a screen displays all choices made as well as the payoffs of each player (including whether their opponent received a fine). Once the pay-

offs in the current round have been displayed, a random draw determines whether another round in the same match will take place (with probability 0.75) and whether the next game is played with or without penalty (with probability 0.5). At the end of each experiment, participants were paid 1 euros for every 100 points earned.

In the T5 treatment, games are played in parallel to identify potential simultaneous spillovers. At the beginning of each round, players are randomly organized into groups of three. Each player is randomly assigned one of three role (Role A, Role B, or Role C). The probability of continuation of each round is determined in the same was as in all other treatments. Each game is independent. The role assigned to each player and the choices made in each round have no direct influence on subsequent rounds.

In each round, Role A must make two choices, one affecting role B and the other affecting role C. This appears as a split screen for for Role A divided vertically into two panel, one corresponding to A's interaction with B (against a blue background), the other with C (against a green background). Roles B and C have only one choice to make and their decisions affect only role A. From the perspective of B and C, interactions occur on a single full screen, against a blue background for Role B and a green background for Role C. All choices made in a give round are made simultaneously (players do not observes the decisions made by the others before choosing).

In each interaction between A and B, the payoffs are equivalent to those of T1 where no legal rule exist. A and C play T2, where legal rules do exist and defection is penalized by $aF = 10$. Consequently, A simultaneously interacts in environments where legal rules are present and where they are absent.

An important aspect of our experiment is that subjects play games repeatedly to allow for learning in both a sequential and simultaneous manner. This is important because of the potential for spillover effects to emerge as subjects are exposed to different environments. The use of both sequential and simultaneous procedures allows us to isolate these potential spillovers across treatments that induce subjects to obey the law because of informal sanctions and the emergence of cooperative norms.

The experimental protocol and programs used to implement in the lab the games described above are innovative with respect to their ability to i) implement a unique matching procedure, particulary in T5 where participants are matched in trios, and *ii*)use a split screen where two games are played in parallel, which to the best of our knowledge has never been done before[3].

Parameters of each treatment

The parameters of the prisoner's dilemma games in this study were chosen following those of Dal Bó and Fréchette [2011]. In their article, Dal Bó and Fréchette [2011] analyze the evolution of cooperation as subjects gain experience, for different probabilities of continuation and payoffs to cooperation. As in Blonski and Spagnolo [2001], they find that support in equilibrium and risk dominance are necessary conditions for cooperation to increase with experience.

Similarly, we examine subjects' behaviour for a variety of parameters: environments where cooperation is not supported in equilibrium (for a given probability of continuation) versus those in which deviation are penalized, effectively bringing cooperation into equilibrium. The results from the legal sanction environment, in which cooperation is both an equilibrium and risk dominant action, are measured against levels of cooperation in games with equivalent payoff functions, but where the sanctions are implicit in the parameters of the game. The specific parameters of our experiments were chosen based on the payoff functions of Dal Bó and Fréchette [2011].

To derive the Base game treatment parameters, we reverse engineer the stage game payoffs of Dal Bó and Fréchette [2011] for which cooperation is a subgame perfect Nash equilibrium and risk dominant action when the probability of continuation $\delta = {}^3/_4$.

[3]Several participants, many of whom were engineers and mathematicians at l'École polytechnique, recognized the innovative T5 design and commented on it to us after the session.

Dal Bó and Fréchette [2011] Stage Game Payoffs

Player 2

Player 1	C	D
C	40, 40	50, 12
D	50, 12	25, 25

We reverse the above stage game payoffs from Dal Bó and Fréchette [2011], for which cooperation is both payoff and risk dominant given a probability of continuation is $\delta = {}^{3}/_{4}$, to calculate the payoffs to deviation that make cooperation neither risk dominant nor payoff dominant. From this we define the parameters of our benchmark treatment absent no legal rules in which defection is optimal. The difference in these payoffs represents the minimum fine aF necessary to ensure that cooperation can be supported as a subgame-perfect equilibrium and risk dominant action[4].

The penalty in our experiment is $F = 10$ with probability $a = 1$ applied to thsuch that deviation is strictly dominated when $\delta = {}^{3}/_{4}$. From this value we derive the parameters of the *Base game* treatment below.

Base game Payoff Matrix

Player 2

Player 1	C	D
C	40, 40	60, 12
D	60, 12	35, 35

[4]The process of reverse engineering the payoff function is detailed in the Appendix.

Under the *Base game* treatment stage game above, cooperation is neither risk dominant nor supported in equilibrium when $\delta = {}^{3}/_{4}$. For all treatments[5] the parameters of the game are those defined in the payoff matrix above. Sanctions in penalty games are imposed externally and do not directly alter the parameters of the games. The parameterizations of each treatment are defined in detail in the Appendix.

[5]With the exception of the Payoff equivalent, which appears exactly as the Dal Bó and Fréchette [2011] matrix above.

If individuals are rational, there is no need to protect them against their own choices.

Daniel Kahneman

5 Theoretical background

THIS EXPERIMENT is designed to study the cooperation-inducing effects of formal institutions and in particular whether exposure to different legal environments has any potential spillover effects. The variables in each treatment were carefully chosen to avoid introducing other confounding influences. For instance, in all games where legal rules apply, penalties to defecting are certain, eliminating complicated issues of risk aversion created by probabilistic outcomes [Kahneman and Tversky, 1979, Tyran and Feld, 2006]. The parameters and procedures of the experiment, allow us to test the the following questions, posed in Chapter 1:

QUESTION 1: *What is the effect of legal rules (legal punishments for choosing the non-cooperative option) on levels of cooperation? More specifically can apparently mild sanctions play a role by changing the equilibrium play in a repeated game setting?*

Standard economic principle-agent theory predicts that formal institutions raise levels of cooperation by imposing higher marginal costs on deviating, or increasing marginal monetary benefits [Frey and Jegen, 2001]. In the control treatment where no legal sanctions exist, the material incentives are such that symmetric defection is the subgame perfect Nash equilibrium and risk dominant strategy. In repeated games with penalties, sanctions on defecting are sufficiently high to achieve bilateral cooperation. Since the private marginal return from defecting is smaller than the long-run incentive to cooperate, cooperation is optimal. Therefore, with respect to Question 1, the standard game-theoretic prediction is for subjects to always deviate (play *D,D*) in T1 and to always cooperate (play *C,C*) in T1.

These predictions assume that all agents are strictly self-interested and fully rational and that this is common knowledge. If however, we assume that one agent is strictly self-interested and fully rational, but believes that his opponent is *conditionally cooperative*, he may have an incentive to cooperate in the initial round and then deviate thereafter. This strategy is motivated by the belief that others interpret previous cooperation as an accurate signal of their partner's type. If a player is not strictly self-interested but conditionally cooperative and interprets previous cooperation as a signal that others are cooperative, then upon observing symmetric cooperation, he may have an incentive to continue engaging in mutually cooperative behaviour.

QUESTION 2: *Do legal punishments have an additional effect than just changing payoffs (extra deterrent effect)?*

Subjective Expected Utility theory predicts no change in behaviour from changing the description of payoffs. Such a change would violate the axiom of "descriptive invariance," which implies that identical options should lead to identical choices, regardless of how the options are framed. In other words, we should observe no significant difference between T2 and T3. This implies that law works only through deterrence, i.e. by changing the tradeoff between the marginal benefits and costs. behaviour is affected only by the severity of sanctions and the probability they will be levied. From this perspective, if material incentives are constant, explicit legal rules should not alter behaviour.

However, according to the psychological theory of framing presented by Kahneman and Tversky [1979], changing the description of otherwise equivalent payoffs may elicit a different psychological reactions that can change behaviour in meaningful ways. Bohnet and Cooter [2003] identify two possible framing effects that redescribing a reduction in payoffs as a "penalty" might invoke. First, "penalty" implies wrongdoing, which might trigger negative psychological costs such as *guilt*. Particularly in public goods games, where the non-cooperative choice can be understood as "harming others." Conversely, the cooperative choice can be described as "benefiting others," inducing a "warm-glow" effect. Second, "penalty" suggests a material *loss*, which may cause loss averse individuals to change their behaviour [Kahneman and Tversky, 1979].

"Framing" is used here to refer to both framing and coordination effects of law. That is, the indirect effect laws by changing both preferences and beliefs. Framing defection as "illegal" may elicit intrinsic motivation to avoid disobeying the law. By extension, laws may change individuals' beliefs about others are likely to behave. In this sense, law offers an equilibrium selection principle around which individuals can coordinate behaviour. If material incentives and psychological framing effects from formal institutions are complementary, subjects should, on average,

cooperate more in T2 than in T3.

However, individuals are often exposed to different legal environments, and may interact simultaneously across jurisdictions in which the same activity received different legal coverage. Unless behaviour is state-dependent, this will likely change the dynamic between the deterrent and psychological effects of legal rules and its affect on behaviour. This reasoning leads us to the following question.

QUESTION 3: *When some areas are covered by legal rules while others are not, are there spillovers from one to the other? In what direction?*

The law is reason, free from passion.

Aristotle

6
Experimental Results

If we wish to make inferences about the causal links between formal institutions on behaviour, theory must be complemented with empirical data from experiments. While theory allows us to make predictions about the role of institutions, experiments allow us to test hypothesized effects against observed behaviour. Examining play in artificial settings helps to expose elements of human psychology and cognition that would ordinarily be unobservable. However, one must bear the artificiality of such experiments in mind when interpreting the results. As Peysakhovich and Rand [2013] rightly emphasize, "experiments are useful because of what they reveal about the psychology produced by the outside world, rather than themselves being a good representation of that world."

In this section we present experimental results on the cooperative behaviour of participants across treatments. We asses whether mild sanctions are effective

mechanisms to achieve cooperation, and if so, whether their influence has an indirect effect on cooperation in environments absent legal rules. That is, whether spillover effects emerge.

6.1 Main Experimental Results

Table 6.1.1: Session Descriptive Data

		Session			
		1	2	3	4
Duo	Treatment	T1	T2	T4	T3
	Date	11/03/2014	13/03/2014	13/03/2014	18/03/2014
	Time*	17:45-18:09	12:30-12:59	17:45-18:22	12:30-13:01
	Subjects	18	18	16	20
	Student†	61.11%	16.67%	81.25%	15%
	Max matches	36	20	26	26
	Ave rounds	4.11	3.42	3.44	4.13
	Ave earning	33.61	16.15	25.76	19.85
	Max earning	41.30	23	34.7	23.12
	Min earning	22.09	12.50	17.74	15.05
Trio	Treatment	T5	T5		
	Date	11/03/2014	18/03/2014		
	Time*	12:30-13:06	17:45-18:26		
	Subjects	18	15		
	Student†	22.22%	60%		
	Max matches	25	29		
	Ave rounds	3.49	3.82		
	Ave earning	20.97	23.67		
	Max earning	32.99	35.74		
	Min earning	7.90	14.24		

* Starting scheduled time and actual final time.
† Percentage of university students in the session.

Six experimental sessions were conducted between March 11 and 18 2014 (see Figure 9.1.1). A total of 105 individuals from l'École Polytechnique including students, faculty and staff participated in the experiment, with an average of 17.5 sub-

jects per session, a maximum of 20 and a minimum of 15[1]. No subjects had prior experience participating in any treatment of our experiment.

The instructions[2] pertaining to the treatment to be played were read silently by each participant, and then repeated aloud by the experimenter[3]. Once prompted, subjects began playing the associated Prisoner's Dilemma game. The parametarization of each stage game is provided in Appendix A. As in Bohnet and Cooter [2003][4] refer to a reduction in payoffs as a "penalty" to imply that the act is wrong, rather than using a reduction in payoffs was referred to explicitly as a "penalty," whereas morally neutral notation, X or Y, was used to denote the actions available to each player, corresponding to the choice of C or D, respectively. Subjects made their choice of action anonymously and simultaneously through computer terminals. These experimental procedures help to minimize influence from additional factors (such as reputation in the case of non-anonymous interactions) that are likely to confound interpretation of the results.

At the end of each round, subjects were informed of their partner's action and their payoff, and whether a penalty had been imposed on a defecting action. Total match payoffs were displayed on the screen at the end of each stage game. Subjects were paid 1 for every 100 points earned throughout all rounds of all games played in the session. The subjects earned an average of 23.34, with a maximum of 41.30 and a minimum of 7.90. The average number of rounds per match was 3.75, and the maximum was 34 rounds.

---

[1]See Table 9.1.1.

[2]The instructions for the *Base game* treatment and the sections from the instructions for the other sessions that are unique to the treatment are included in Appendix B.

[3]The author.

[4]The choice of language used in an experiment must be made with great care, as even the slightest nuance in terminology can unintentionally convey information that may impact results. For example, Bohnet and Cooter [2003] note that, "'payoff' describes the sanctions effect in morally neutral language, whereas 'penalty' implies that the act is wrong."

Table 6.1.2: Framing effects: rate of individual cooperation

	All repeated games				
	Base Game		Penalty		Payoff Equivalent
First Round	11.01	$<^{***}$	32.39	$>^{**}$	25.68
	$\vee^{***}$		$\vee^{***}$		$\vee^{***}$
All rounds	6.40	$<^{***}$	27.61	$>^{***}$	17.31
	First repeated game				
	Base Game		Penalty		Payoff Equivalent
First Round	55.56	$=$	55.56	$<$	60.00
	$\vee^{***}$		$\vee$		$\vee^{**}$
All rounds	26.67	$<^{**}$	50.00	$>$	45.31

*, **, *** indicate significance at the 10, 5, and 1 percent level, respectively.

6.1.1 General Description of behaviour

Comparing behaviour across the first three treatments allows us to analyze the effectiveness of formal sanctions in achieving cooperation (i.e. by comparing *Penalty* versus *Base game* treatment), and to isolate potential framing effects (i.e. *Penalty* versus *Payoff equivalent* treatment). The *Base game* treatment only permits one equilibrium, i.e. mutual defection. Given the parameters of the game, informal enforcement mechanisms via repetition fail to shift players from the inefficient all defect equilibrium. The *Penalty* and *Payoff equivalent* treatments investigate how formal and informal mechanisms achieve cooperation, differing only with respect to the framing of payoffs. Under the *Penalty* treatment formal sanctions lower the relative payoff to defection so that mutual cooperation is optimal. Whereas the *Payoff equivalent* treatment changes the *Base game* payoffs directly. The difference can be summarized as an implicit *change* to payoffs in the former versus and an implicit *penalty* to defection in the latter.

Table 6.1.2 highlights potential framing effects of legal rules in our experiment. The top panel shows cooperation rates for all games within the first three treatments; the top row for the first round and the second for all rounds[5] Cooperation

[5] Throughout the paper (unless otherwise indicated) statistical significant is tested using probit

Table 6.1.3: Influence of sanctions: rate of individual cooperation

	No Penalty					Penalty				
All repeated games										
	Base Game		Trio		Random - Duo	Penalty		Trio		Random - Duo
First Round	11.01	<***	23.27	<**	30.38	32.39	<***	44.97	<***	61.93
	∨***		∨***		∨***	∨***		∨***		∨**
All rounds	6.40	<***	16.64	<**	20.58	27.61	<***	34.59	<***	57.26
First repeated game										
	Base Game		Trio		Random - Duo	Penalty		Trio		Random - Duo
First Round	55.56	>	42.42	>	31.25	55.56	>	48.48	<**	68.75
	∨***		∨		∨**	∨		∨		∧***
All rounds	26.67	<**	41.23	>***	19.61	50.00	<	54.54	<***	81.54

*, **, *** indicate significance at the 10, 5, and 1 percent level, respectively.

is clearly higher when it can be supported in equilibrium and is the risk dominant action ($p-$values <0.001 for both first rounds and all rounds). Furthermore, the presence of a penalty has a salient effect on cooperation, evidenced by significantly greater cooperation under the *Penalty* treatment than the *Payoff Equivalent* treatment. This suggests a significant framing function served by the presence of the penalty, inducing more cooperation than when the it is implicit in the parameters of the game.

The bottom panel of Table 6.1.2 shows cooperation rates for the first game, again, with the first and second rows displaying cooperation in the first round and all rounds, respectively. No significant framing effect is observed for the first round. Unlike Dal Bó and Fréchette [2011], we do not observe an increase in cooperation with experience for the first three treatments. Cooperation is greater in the first round than over all subsequent rounds[6]. Duffy and Ochs [2009] observe a similar phenomenon, the so-called 'restart' effect, which they interpret as the "repeated effort by just a few subjects to encourage a social norm of cooperation," which diminishes as additional games are played. [Duffy and Ochs, 2009: 791].

analysis. Standard errors are clustered at the individual level.

[6]This is consistent with equilibrium behaviour in the *Base game* treatment, where defecting is

Table 6.1.3 shows cooperation rates in games played with and without penalties across varying legal landscapes. The left panel presents cooperation of the *Base game* treatment to *Trio* and *Random - Duo* games played without penalty[7]. The probability of a penalty under the *Base game* treatment is zero. Under the *Random - Duo* treatment, the probability that any given match is played with penalties is randomly determined (*Pr*(penalty game)=0.5). Under the *Random - Duo* treatment, players face a 1 in 3 chance of being assigned either Role A, Role B, or Role C in any given match[8]. The outcome of the random draws before each match in the *Random - Duo* and *Trio* treatments are independent of one another.

The payoff functions of penalty games across all treatments are equivalent. As are the payoff functions for all no-penalty games. However, we see that under treatments where the legal environment is constant (*Penalty* and *Base game*) overall cooperation rates tend to be significantly less than when subjects are exposed to changing legal rules (i.e. in the*Trio* and *Random - Duo* treatments). This suggests a potential spillover effect from previous matches played with penalty to those in which penalties are absent[9]. Cooperation is significantly greater in the *Random - Duo* games than in the *Trio* games (significant at the 5 percent level for both first rounds and all rounds). We can speculate that this is the result Role A's behaviour, whose simultaneously interactions in games with and without penalties appear to temper the positive spillover effect. In other words, indirect influence from one legal rules environment to the other may occur in both directions.

The bottom panel shows cooperation in the first game subjects play. The first round of the first game tends to show no significant difference in the rate of cooperation across treatments (except in *Trio* versus *Random - Duo* penalty games). This is to be expected since players have not yet had the opportunity to update their beliefs. Again, cooperation tends to decline as additional games are played, with the exception of *Random - Duo* penalty games. It is interesting that under the *Trio*

---

the payoff- and risk dominant- action.

[7]*Trio* games played between Role A and Role B

[8]The probability of playing in a *Trio* game between Role A and Role B or between Role A and Role C is equal to 0.5 at any given time.

[9]Or from parallel matches with penalty played simultaneously, in the case of Role A.

treatment cooperation in the first penalty games is greater than in either the *Base game* or the *Random - Duo*, significant and the 5 and 1 percent levels, respectively. This may reflect and initial acceleration in learning afforded by the simultaneous interactions of Role A. Whereas under the *Random - Duo* treatment, learning from exposure to different legal rules environments occurs only sequentially.

6.1.2 Evolution of cooperation

To see the effect of learning on players' behaviour, we study the evolution of cooperation under treatments in which cooperation cannot be supported in equilibrium versus those in which it can be supported. The first column of Table 6.1.4 shows the rate of cooperation in all rounds of each match in the *Base Game* treatment, with matches aggregated into sequential groups of ten. To compare inexperienced versus experienced play, we can compare behaviour in the first ten matches with subsequent groups of matches.

From Table 6.1.4 we see that as players gain experience from games in which it is optimal to defect, cooperation significantly declines. Focusing attention to the first eight series of matches, for which we have data on all No Penalty games, cooperation drops by an average of 54.04% between the matches 1-10 and matches 71-80.

While our sample also shows cooperation decreasing with experience when it *can* be supported in equilibrium[10], the decline is much less pronounced. The first three columns of the right panel under Penalty games show an average decline in cooperation of 17.7% between the first and last group of matches for which we have data on all Penalty games.

Interestingly, cooperation under the *Payoff equivalent* treatment drops significantly faster than it does under treatments in which the penalty is explicitly stated, despite equivalent material payoffs.

[10]Cooperation is also the risk dominant action in these treatments.

Table 6.1.4: Percentage of cooperation in Penalty and No Penalty games (all rounds)

All rounds in repeated game	No Penalty Games			Penalty Games			Payoff equivalent
	Base Game	Random	Trio	Penalty	Random	Trio	
1-10	24.24	17.05	33.56	50.00	61.46	44.23	45.31
11-20	12.96	38.33	16.67	50.00	75.81	40.96	20.00
21-30	7.50	37.80	16.67	53.13	50.00	38.89	28.79
31-40	13.95	17.95	31.55	41.23	68.57	37.50	18.06
41-50	4.69	23.26	9.57	32.43	46.05	22.67	26.74
51-60	6.45	14.71	5.70	53.57	68.42	26.52	11.21
61-70	6.14	7.35	7.25	16.67	38.00	38.10	15.48
71-80	6.94	5.17	10.81	18.63	58.57	22.22	16.07
81-90	6.38		8.33	34.62	50.00	27.27	11.54
91-100	7.14			8.33			13.46
101-110	13.79			8.62			11.11
111-120	4.84			20.00			7.14
121-130	7.32			5.56			38.16
131-140	0.00			10.00			5.56
141-150	0.00			25.00			13.24
151-160	0.00						20.59
161-170	1.82						10.53
171-180	0.00						24.19
181-190	12.16						16.67
191-200	1.16						

Note : To estimate the evolution of cooperation with experience, repeated games are aggregated by groups of ten consecutive matches. The total number of groups varies across treatments since different sessions may result in a different number of total repeated games played.

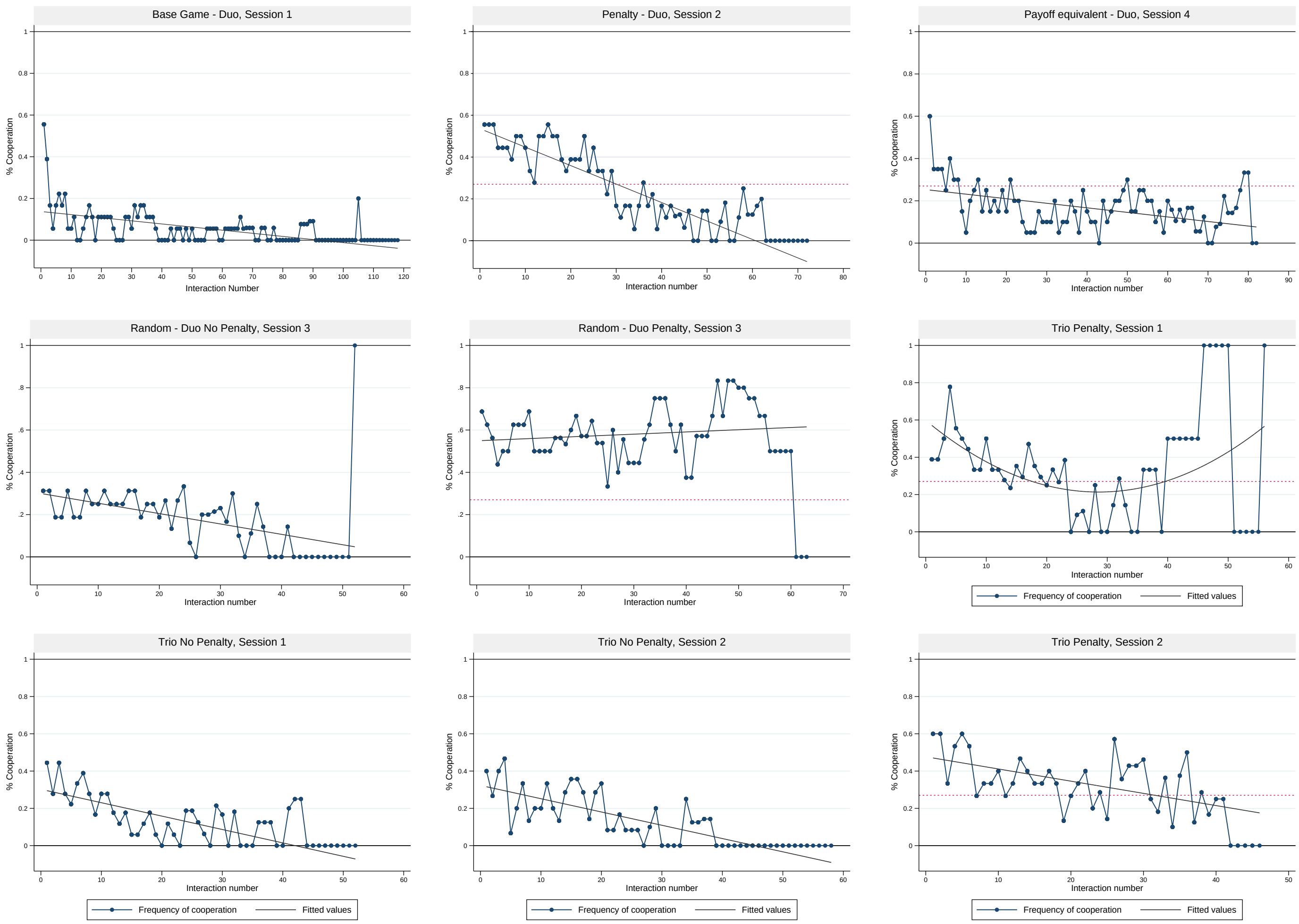

Figure 6.1.1: FREQUENCY OF COOPERATION BY TREATMENT AND SESSION

Note: Cooperation rates aggregated by interaction. The horizontal dotted lines denote the limit of the basis of attraction of *always defect* versus *grim trigger* in games with penalty or *Payoff equivalent*.

Figure 6.1.1 presents data on the frequency of cooperation across interactions in each treatment and game type[11]. When deciding whether it is optimal to pursue a cooperative strategy, such as *grim trigger* (G), or a defecting strategy, such as *always defect* (AD), a player considers both the fundamental parameters of the game and his or her beliefs about the expected probability that his or her opponent will

[11]The term "interaction" is used to denote each decision stage regardless of the match. While "round" is used to specify the decision stage *within* a give match

cooperate or play AD[12]. In Figure 6.1.1, the horizontal dotted line represents the belief of the partner's propensity to play the cooperative strategy that leaves the subject indifferent between the cooperative and defecting strategies for in penalty games[13]. In no-penalty games, it is never optimal to cooperate.

6.2 Empirical analysis

In this section we analyze the experimental data to determine if there is statistical evidence of the effect of legal rules on the probability of cooperation, holding constant participant characteristics that might influence an individual's propensity to cooperate.

The baseline probit specification is

$$P(Cooperation = 1|x_j, \tilde{\mathbf{x}}) = \Phi(\beta_j x_j + \tilde{\mathbf{x}}\tilde{\beta}), \tag{6.1}$$

where subscripts for observations are dropped for simplicity, *Cooperation* is the binary dependent variable (equals 1 if the subject chooses C, 0 otherwise), Φ is the cumulative standard normal distribution function, x_j denote the treatment dummy variables, β_j are the associated coefficients,and $\tilde{\mathbf{x}}\tilde{\beta}$ denotes the linear combination of all remaining explanatory variables and coefficients, including player and game characteristics.

The independent variables included in the baseline probit models in the empirical analysis of cooperative behaviour are described in Table 9.1.1 of Appendix B. The first four variables are individual player characteristics that may affect gameplay behaviour[14]. The first of these is a binary control for gender, which takes the value of one if the player is male, 0 if the player is female[15]; the second is a dummy, indicating whether the subject is a student; the third controls for the participant's

[12] G is never a best response for No Penalty treatments.

[13] The proof is provided in Appendix B.

[14] Data on player characteristics were collected from surveys completed by each participant immediately following the experiment

[15] The author recognizes that a binary measure of gender may be overly reductionist, but this should not have any discernible impact on the results of the present study.

age. The fourth is a discrete variable that measures a player's risk aversion on a scale from 0 ("highly averse to risk") to 10 ("prone to risk taking").

The strategy *Always Defect* (AD) represents the (riskless) prospect, which yields payoff (*D,D*) with certainty. The payoff parameters in our experiment were chosen so that all games in which cooperation is the subgame perfect Nash equilibrium, it is also the risk dominant action. Payoff functions are derived from Dal Bó and Fréchette [2011] according to the adapted definition of risk dominance for infinitely repeated games by Blonski and Spagnolo [2001][16]. As such, cooperation is the optimal choice for even risk averse players. That is, the long-term incentive to cooperate exceeds the short-term incentive to defect when both the "sucker payoff" from choosing to cooperation when the player's opponent defects and the payoff from mutually defecting are taken into account.

Partner cooperated in round 1 of previous match is a binary variable that takes the value of one if the players opponent cooperated in the first round of the last match, and zero if he or she defected. *Number of rounds in the previous match* is a discrete variable, indicating the total number of rounds realized in the last match the subject played. Dal Bó and Fréchette [2011] suggest that the behaviour of partners and the number of rounds in the previous game are positively correlated with the probability that player *i* will cooperate in the current match. In other words, a subject matched with someone who played *C* in the first round is more likely to attempt to cooperate in the following match, than if the opponent had defected. They interpret this phenomena as evidence of updating behaviour; "if subjects have an estimate of the fraction of the population playing a strategy that starts by cooperating they update their estimate upward after observing a sample consistent with that strategy" (Dal Bó and Fréchette, 2011: 422). This effect is compounded positively by the length of a match. The longer the sustained cooperation, the more likely the player is to cooperate in subsequent matches.

We additionally include a binary variable, *subject cooperated in round 1 of match*

[16]Blonski and Spagnolo [2001] derive a simplified definition that can be applied to infinitely repeated games, based on the concept of risk dominance in one-shot games introduced by Harsanyi and Selten [1988]. A detailed description of their method is given in Chapter 3.

1, which equals 1 if the subject played *C* in the first round of the first match, 0 otherwise. We predict that this variable will be positively related to cooperation, as it reflects players' initial propensity to cooperate. In these games, 55.6% of the participants cooperated in the first round played.

6.2.1 Test for framing effects

Table 6.2.1 examines potential framing effects resulting from explicitly stated legal rules on the probability of cooperation. All equations are estimated by probit regression, with standard errors clustered at the individual level. The first three columns include data on"all rounds," while the sample in the last column is restricted to game play in the "first round" of each match. The base specification, reported in column (1) estimates the probability of cooperation, conditional on the individual characteristics described above. Table ref{framereg also reports the difference in the marginal effects for the two treatment dummy variables, with all other independent variables taken at their means. [17].

The treatment dummies in Table ref{framereg equal one if the game played was under either the *Payoff equivalent* or the *Penalty* treatment, zero otherwise (base level specified as *Base game* treatment).

Model (1) presents the probability of cooperation across the first three treatments, holding constant individual player characteristics.

To investigate the potential framing effect of legal rules, column (2) introduces treatment dummy variables for (*Penalty* or *Payoff equivalent* with *Base game* as the base treatment. The coefficients on the treatment variables are positive and significant at the 1 percent level, indicating that cooperation rates under the *Penalty* and *Payoff equivalent* treatments are significantly higher than in the *Base game*. This suggests cooperating increases not only when it can be supported in equilibrium, as in the *Payoff equivalent* treatment, but also when the payoff function is implicitly altered by the presence of an externally imposed penalty. The legal rules can

[17]The marginal effects show how the $P(\text{Cooperation}=1)$ changes as the treatment variable changes from 0 to 1, holding all other variables at the means).

Table 6.2.1: FRAMING EFFECTS PROBIT REGRESSIONS

	All rounds			First round
$P(\text{Cooperate}_i{=}1)$	(1)	(2)	(3)	(4)
Gender	-0.33	-0.11	-0.08	-0.05
	(0.20)	(0.19)	(0.19)	(0.23)
Student	0.13	-0.37	-0.10	0.21
	(0.27)	(0.38)	(0.36)	(0.47)
Age	0.01	0.00	0.00	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Risk aversion	-0.01	0.07^{**}	0.06^{*}	0.06
	(0.04)	(0.03)	(0.03)	(0.04)
Penalty treatment		0.80^{***}	0.85^{***}	0.65^{***}
		(0.17)	(0.18)	(0.25)
Payoff equivalent treatment		0.53^{**}	0.59^{**}	0.74^{**}
		(0.26)	(0.27)	(0.36)
Partner cooperated in round 1 of previous match			0.23^{**}	0.33^{**}
			(0.11)	(0.16)
Number of rounds in previous match			-0.03^{***}	-0.07^{***}
			(0.01)	(0.02)
Subject cooperated in round 1 of match 1			0.42^{**}	0.72^{***}
			(0.19)	(0.23)
Observations	5312	4118	4118	1068

Notes: Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

promote prosocial cooperation.

Holding all additional independent variables at their means, interacting in a *Penalty* game increases the probability of cooperation by 17.28% relative to the *Base game* treatment. Interestingly, the marginal effect of the *Payoff equivalent* treatment on the likelihood of cooperating is only 9.87%, a 7.41 percentage point differential (significant at the 10 percent level). In other words, an explicit penalty increases the predicted probability of cooperation 7.41 percentage points more than when the penalty is implicit in the payoff function of the game. This suggests that there is indeed a salient framing effect created by the presence of legal rules in our experiment.

Columns (3) modifies column (2) by including additional game-play dynamics described above. These additional controls do not change the estimated coefficients of Model (2) in an important way. The marginal framing effect differential is 7.32 percentage points, significant at the 10 percent level. That is, the probability of cooperating is 7.32 percentage points higher under the *Penalty* treatment than under the *Payoff equivalent*, all other variables taken at their means. The positive and highly significant coefficient on *Subject cooperated in round 1 match 1* indicates that, all else equal, the predicted probability that a subject cooperates is significantly greater if he or she cooperated in the first round of the experiment.

Column (4) differs only with respect to the sample, which is restricted to play in the first round of each match. Again, whether or not the subject cooperated in the first round of the first match has a positive and statistically significant effect on the overall probability of cooperation. If a subject cooperates from the outset, he or she is more likely to cooperate at the start of subsequent matches. While the effect of both the *Payoff equivalent* and the *Penalty* treatment remain statistically significant, we no longer observe a salient framing effect created by an explicit penalty[18].

Framing effects in punishment behaviour

Table 6.2.2 examines the probability that a subject defects as a function of their partner's behaviour in previous rounds in the *Payoff equivalent* versus *Penalty* treatment. In other words, whether punishment behaviour is driven by framing effects.

As before, column (1) present the baseline model, controlling only for individual characteristics, and column (2) introduces the effect of treatment type. Column (2) perfectly mirrors its Table 6.2.1 counterpart. Subjects are less likely to defect under the *Penalty* treatment than the *Payoff equivalent*. This may reflect the psychological framing effect of the penalty, leading players to coordinate behaviour on the Pareto-superior cooperative equilibrium.

Columns (3) and (4) examine the propensity of players to resort to punish-

[18] The difference in the marginal effect on the predicted probability of first round cooperation for *Penalty* versus *Payoff equivalent* is not significantly different from zero.

Table 6.2.2: Punishment Framing Effects Probit Regressions

	All rounds			First round
$P(\text{Defect}_i=1)$	(1)	(2)	(3)	(4)
Gender	0.33	0.11	0.09	0.08
	(0.20)	(0.19)	(0.18)	(0.23)
Student	-0.13	0.37	0.12	-0.20
	(0.27)	(0.38)	(0.37)	(0.47)
Age	-0.01	-0.00	-0.00	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Risk aversion	0.01	-0.07^{**}	-0.06^{*}	-0.06
	(0.04)	(0.03)	(0.03)	(0.04)
Penalty treatment		-0.80^{***}	-0.88^{***}	-0.17
		(0.17)	(0.31)	(0.41)
Payoff equivalent treatment		-0.53^{**}	-0.22	0.06
		(0.26)	(0.35)	(0.50)
Partner defected in round 1 of previous match			0.39^{*}	0.94^{***}
			(0.21)	(0.26)
Number of rounds in previous match			0.03^{***}	0.07^{***}
			(0.01)	(0.02)
Subject cooperated in round 1 match 1			-0.42^{**}	-0.74^{***}
			(0.19)	(0.23)
IT: *Penalty* × *Partner defected rd 1 prev. match*			0.06	-0.59
			(0.29)	(0.38)
IT: *Payoff eq.* × *Partner defected rd 1 prev. match*			-0.48^{*}	-1.01^{***}
			(0.26)	(0.34)
Observations	5312	4118	4118	1068

Notes: Clustered standard errors in parentheses. $^{*}\ p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

ment when faced with non-cooperative partners in the past. As in Table 6.2.1, probit model (4) is restricted to interactions that take place in the first round of each match. Game-dynamic controls are introduced, as well as two interaction terms that allow the effect on the probability of defection of former opponents' behaviour to depend on treatment type. More specifically, the interaction terms capture the difference in the effect of a player's opponent defecting in the first round of the previous match for *Penalty* versus *Base game* treatments and for *Payoff equiv-*

alent versus *Base game* treatments, respectively.

The positive coefficient on *Partner defected in round 1 of previous match* implies that for *Base games*, the effect of having a non-cooperative partner in previous rounds increases the probability that a subject will defect in the future. This effect is significant at the 10 percent level. When we look at how it depends on treatment type, we learn that the predictive power of the *Payoff equivalent* dummy is conditional upon its interaction with *Partner defected in round 1 of previous match*. The negative coefficient on the interaction term (significant at the 10 percent level) suggests that subjects who faced non-cooperative partners in the first round of the last game are less likely to defect under the *Payoff equivalent* treatment than the *Base game* treatment.

The negative effect of the *Penalty* dummy implies that subjects are significantly less likely to defect under the *Penalty* treatment versus the *Base game* ($p-$values $<$ 0.001). Interestingly, the interaction term for the *Penalty* treatment is positive, suggesting that players are more likely to defect under the *Penalty* versus the *Base game* treatment if paired with non-cooperative partners in the past are induce players. This effect could be interpreted as attempts to enforce cooperative norms by punishing non-cooperative behaviour. The coefficient on the interaction term is not significantly different from zero. However, in the nonlinear probit, the magnitude and statistical significance of the interaction effect can vary greatly (see Figures 6.1.1 and 6.1.2).

Figure 6.1.1 shows how the interaction effects and their significance depend on other covariates. The top graphs plot the interaction effects against predicted probabilities for the Penalty treatment on the left, Payoff equivalent on the right. The full interaction effect plotted in blue is the cross-partial derivative of the expected value of the dependent variable, $Defect_i$. The red dashed line is the interaction effect calculated by the conventional linear method.

The bottom graphs test the statistical significance of the interaction effect, based on the estimated cross-partial derivative. Here, z-statistics of the interaction effect

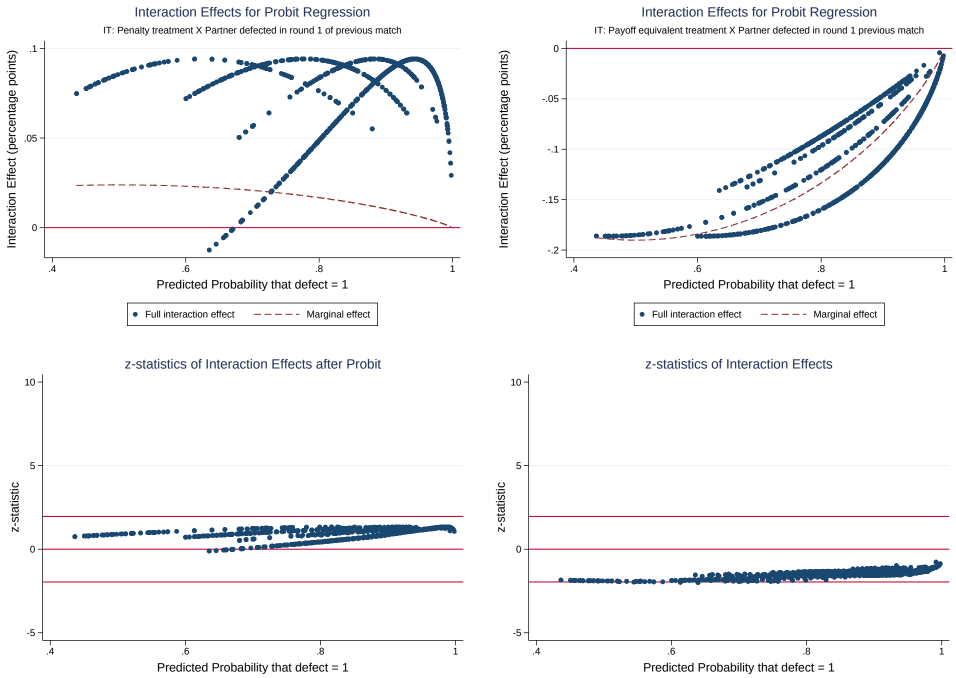

Figure 6.2.1: Framing Effects in Punishment Behaviour - All rounds

are plotted against predicted probabilities on the x-axis.

The top right graph of Figure 6.1.1 confirms that if a player's opponent defected in the first round of the previous match, the subject is less likely to respond by defecting in the current round under the *Payoff equivalent* treatment than under the Base game treatment. For subjects whose predicted probability of defection is very high, the difference in punishment behaviour across the treatments diminishes. The full interaction effect is statistically significant for several observations at $P(Defect_i = 1) \approx 0.6$ (for full interaction effect, z-statistic$\in$ $(-1.991, -0.770)$). The left panel also confirms the positive interaction term for the *Penalty* treatment reported in Table 6.2.2. This implies evidence of the expressive function of legal rules, which amplifies the psychological effect opponents' behaviour in the past on the subject's behaviour in the present.

Focusing on punishment behaviour in the first round of each match, we see the

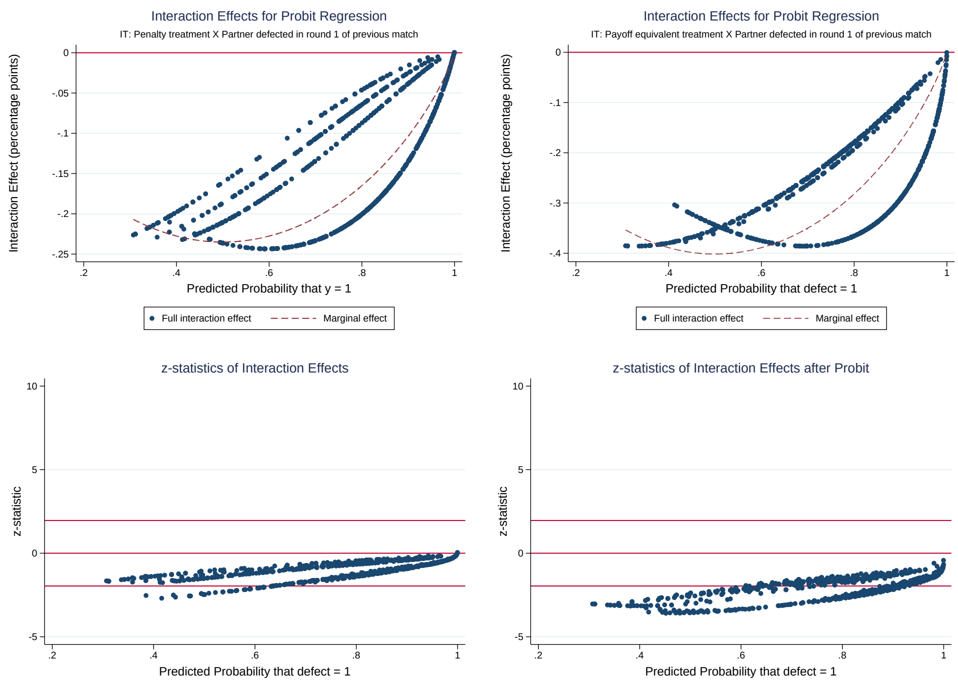

Figure 6.2.2: Framing Effects in Punishment Behaviour - First round

immediate reaction to non-cooperative behaviour in the previous match. Here, the framing effect between Penalty and Payoff equivalent treatments seems to disappear. Subjects under both treatments are less likely to defect if their last partner was non-cooperative than subjects under the Base game treatment. Despite the lack of significance of the coefficient on the Penalty interaction term in column (4) of Table 6.2.2, the full interaction effect for both treatments is statistically significant for many observations (see the bottom panel of Figure 6.1.2).

The shape of the Payoff equivalent interaction effect in the first rounds is quite similar to that in all rounds. On the other hand, the Penalty interaction effect changes dramatically across all rounds. This dynamic might be explained by what Bohnet et al. [2008] refer to as "betrayal aversion." That is, the aversion to take a risk (to attempt to cooperate in our case) when decisions by other individuals are the main source of uncertainty. Rational choice theory does not allow for this dis-

tinction with respect to the source of the risk. However, Bohnet et al. [2008] find that in situations where the agent of uncertainty is another person, notions such as trust and trustworthiness matter. Traditional definitions of risk aversion alone may not account for people's behaviour under uncertainty.

From this perspective, even when cooperation is both payoff and risk dominant, the explicit signal created by the penalty of being *wronged* by an opponent who defects, makes the subject more inclined to punish non-cooperative behaviour. Over the course of a given match the framing effect becomes more salient. The relative decline in the predicted probability of defect at the beginning of the next match is evidence of a "restart" effect. Subjects are willing to retry to establish cooperation with a new partner.

Sequential spillovers

The *Random - Duo* treatment is designed to mimic situations where institutional spillovers occur sequentially. Spillovers might be temporally sequential, deriving from changing legal rules in a given jurisdiction (e.g. the ratification of a new law; a political regime shift resulting in more or less strict legal rules; implementation of legal transplant). Spillovers might also be spatially sequential. In this situation, an agent travels from country X, where the legal rules surrounding a particular action are relatively stringent, to country Y, where legal rules are relatively lax, or vice versa. The agent's experience under the legal rules of country X may spillover onto his or her behaviour in Y. To illustrate, imagine again the anti-litter law example, and suppose that an individual moves from Singapore, where anti-litter laws are relatively strict, to Paris, where they are relatively less strict. If behaviour is state-dependent then the subject should be more inclined to litter in Paris where there is little incentive not to do so. If however, institutional spillovers exist, the subject's prior experience in Singapore may reduce his or her propensity to litter elsewhere.

To study how exposure to different legal environments in the past affects cooperative behaviour today, we analyze play under the *Random - Duo* treatment. Under this treatment, as previously discussed, players interact in a series of Penalty

and Base games, referred to as "Game with Penalty" and "Game without Penalty" in the instructions of the treatment, respectively. The probability of playing either game type in any given match, and thus sequence of games, was determined randomly ($P(\text{Penalty game}) = P(\text{No penalty game}) = 0.5$).

Rational choice theory predicts that actions taken in a given match are independent of those taken in previous matches. According to this view, penalties influence behaviour only by changing the relative payoffs to cooperation and defection. However, experimental evidence shows that as subjects gain experience they may reassess the gains of attempting to establish a cooperative relationship with a partner, and they may modify their behaviour accordingly[19] [Dal Bó and Fréchette, 2011]. As players interact with different partners, they can more accurately estimate the actual proportion cooperative players in the population, and thus the probability of achieving successful cooperation.

This process of updating beliefs about the distribution of "types" (e.g. prosocial; conditionally cooperative; rational; opportunistic) depends on the features of the game affecting incentives. Under the *Random - Duo* treatment, players learn from signals created by the legal rules environment surrounding a particular action. For example, if a player cooperates in a no-penalty game, this conveys more information about his or her type than in a penalty game where the player has a monetary incentive to do so. The penalty diminishes the potency of the prosocial signal. Conversely, if a player defects in a penalty game where cooperation is optimal, this sends a strong non-cooperative signal. Defection in a no-penalty game indicates only that the subject is behaving rationally. Thus, the changing legal landscape allows players to develop a more accurate perception of the distribution of cooperative types in the population, and thus the probability of establishing successful cooperation.

However, signals may be altered if there is a psychological effects created by the penalty itself. Penalties may foster the emergence of cooperative norms that lead to higher overall cooperation rates. In this case, the decision to cooperate may be

[19]The process of learning in the presence of formal sanctions is developed theoretically in the Appendix.

less responsive to opponents' past behaviour.

A striking feature of play under the *Random - Duo* treatment is its consistency with sub-game perfect Nash equilibrium behaviour in both penalty and non-penalty games (see Figure 6.1.1). In the penalty games we observe cooperate rates climbing above 80%, remaining well above the basin of attraction of defection (marked by the horizontal red dotted line)[20]. In no-penalty games cooperation rates decline to near-zero levels with experience, as theory predicts.

This begs the question why cooperation rates are so much higher in the penalty games under the *Random - Duo* treatment than those in either the *Penalty* or *Payoff equivalent* treatments. The uncertainty surrounding the particular legal environment subjects are exposed to must influence players' strategies. In other words, effects of past experience in penalty and no-penalty games spillover onto decisions made in subsequent interactions.

Table 6.2.3 investigates whether spillovers from previous rounds under the *Random - Duo* treatment have significant effects on current behaviour. Columns (1) and (3) look at the predicted probability of cooperation in early games, split at the median 10th match of the sample. Columns (2) and (4) study behaviour in later games after the 10th match. As before, standard errors are clustered at the individual level.

Columns (1) and (2) are our baseline probit regressions[21], additionally controlling for game type (*Random game with penalty*) and whether the previous match was a penalty game (*Penalty in previous match*). We immediately see that penalty games have dramatically higher cooperation rates than no-penalty games, significant at the 1 percent level. The marginal effect on the predicted probability of cooperation of a discrete change from no-penalty to penalty game is a 39.72 and 36.6 percentage point increase (significant at the 1 percent level) in early games and in late games, respectively. What is noteworthy about columns (1) and (2) are the coefficients on *Penalty in previous match*. The negative coefficient in col-

[20]*See* Appendix B for a detailed explanation and derivation of the basin of attraction.

[21]Individual player characteristics not reported.

Table 6.2.3: Sequential Spillovers Probit Regressions

	Early games	Late games	Early games	Late games
$P(\text{Cooperate}_i{=}1)$	(1)	(2)	(3)	(4)
Random game with penalty	1.04^{***}	1.34^{***}	1.04^{***}	1.35^{***}
	(0.12)	(0.15)	(0.13)	(0.15)
Partner cooperated in round 1 of previous match	0.49^{***}	0.34^{**}		
	(0.14)	(0.17)		
Number of rounds in previous match	0.01	0.02	0.00	0.01
	(0.02)	(0.03)	(0.02)	(0.03)
Subject cooperated in round 1 of match 1	1.11^{***}	0.35	1.01^{**}	0.35
	(0.41)	(0.45)	(0.43)	(0.43)
Penalty in previous match	-0.38^{***}	0.48^{***}		
	(0.11)	(0.14)		
IT: *Penalty in previous match*			-0.11	0.64^{***}
× *Partner cooperated in round 1*			(0.14)	(0.16)
IT: *Penalty in previous match*			-0.61^{***}	0.10
× *Partner defected in round 1*			(0.18)	(0.25)
IT: *No penalty in previous match*			0.47^{**}	0.15
× *Partner cooperated in round 1*			(0.20)	(0.24)
Observations	655	501	655	501

Notes: Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

umn (1) implies that, in early matches, if the previous match was a penalty game, players are *less* likely to cooperate in the current match. In later rounds, however, if the last match was a penalty game, players are *more* likely to cooperate.

Columns (3) and (4) offer a deeper analysis of the previous-match dynamics. By interaction the *Penalty in previous match* binary variable with the probability that the subject's partner cooperated in round 1 of the previous match, we introduce three interaction terms, with the base level specified as the interaction between no penalty in the previous match and the subject's partner did not cooperate in round one. Comparing the dynamics of these interaction effects, we can isolate the various factors that might be influencing game-play behaviour.

In columns (3) and (4) the three interaction terms (henceforth IT1, IT2, and IT3, respectively) distinguish learning from spillovers of psychological effects from

previous matches. The interaction terms are defined as follows: IT1 captures the interaction between the presence of a penalty in previous match and the partner having cooperated in the first round; IT2 is the effect of a penalty and the partner having defected; IT3 is the effect of no penalty in the previous match and the partner cooperating in round 1.

In the absence of spillover effects, i.e. if behaviour is motivated only by material incentives and the updating of beliefs, the effect of encountering a cooperative partner in the previous match is greater in no-penalty than penalty games. If a partner cooperates in a penalty game, the material incentive created by the penalty reduces the signal about the partner's type, preventing players from updating beliefs. Cooperation in games where legal rules are absent, however, improves perceptions about the overall cooperativeness of the population. Conversely, defecting in games where legal rules are present signals more information about the proportion of non-cooperative types than in no-penalty games, where defection is optimal.

This is largely what we observe in early matches of the *Duo - Random* treatment. The information conveyed by the interaction terms in column (3) suggests that learning dominates the psychological effect of past institutional environments. This is reflected by the positive and significant effect of IT3, and the wide gap between IT1 and IT3; cooperation in the past significantly increases the probability of cooperation in the current match if it occurred in a no-penalty game, thereby allowing players to make inferences about the distribution of types. Whereas in penalty games, the inability to update beliefs yields the effect of cooperation in the previous match statistically insignificant. The negative and highly significant coefficient on IT2 is further evidence of a dominant learning effect in early games. Players are more likely to punish in response to non-cooperative past encounters if they occurred in penalty games, as opposed to no-penalty games where learning is weaker.

In later rounds, we see that the effect of learning diminishes and the psychological spillover from past institutions dominates behaviour. The psychological effect of the penalty itself boosts cooperation in IT1 and IT2, and we find a positive and

significant spillover effect. This is confirmed by the loss of magnitude and significance of the learning effects revealed by IT2 and IT3.

To summarize, we can think of beliefs about the distribution of types as a learning curve, which is initially very steep and then levels off over time. Thus, learning dominates players' strategies in early games. With experience, however, the information conveyed by the actions of former partners has a marginally declining effect on the updating of beliefs. As learning levels off, the psychological effects created by legal rules, such as the framing effects discussed in the previous section, start to influence behaviour. While rational choice theory assumes independence of each sequential game, what we observe in the lab suggests that the institutional environments experienced in the past have important spillover effects on behaviour in the present.

Simultaneous spillovers

The *Trio* treatment is designed to reflect institutional spillovers that occur simultaneously. Spillovers of this type may result from situations in which an agent interacts in parallel in two jurisdictions (e.g. contractual relationships that occur internationally, across countries with divergent contract laws) or over two different tasks covered by different legal rules (e.g. two contractual relationships within the same jurisdiction, one enforced by stringent contract law, the other absent formal coverage). Simultaneous spillovers might also occur in environments where multiple legal systems coexist simultaneously (i.e. under conditions of legal pluralism).

The innovative experimental protocol and unique matching procedure used to implement this treatment in the lab allow for learning that occurs in a simultaneous manner. In this treatment, participants are matched in trios and randomly assigned one of three roles, namely: Role A, Role B, and Role C. Within each trio, two repeated games are played in parallel: a no-penalty game between Role A and Role B and a penalty game between Role A and Role C. Role A interacts simultaneously in a vertically divided screen. On one side of the screen, labelled "Game without

penalty," A interacts with B, on the other side, labeled "Game with penalty" A interacts with C. B and C interact exclusively with Role A on a single-panel screen, labeled either "Game without penalty" or "Game with penalty".

Comparing the simultaneous procedure to the sequential design of the T4 allows us to isolate potential spillovers resulting from interacting simultaneously in different legal environments. That is, whether Role A's propensity to cooperate is impacted by simultaneous exposure to both penalty and no-penalty games. Additionally, we examine whether C's behaviour against A indirectly influences the probability that B and A cooperate, and vice versa. Presumably, if C is cooperative, A is more likely to reciprocate by cooperating. Successful cooperation between C and A may influence A's interaction with B, indirectly increasing the probability of cooperation.

Table 6.2.4 presents cooperation rates for each role under the *Trio* treatment in either no-penalty or penalty games. Cooperation rates under the *Base game* and *Penalty* treatments provide benchmarks to highlight potential spillovers under the *Trio* treatment. As in Table 6.1.2, the top panel shows cooperation rates for all repeated games, with no-penalty games on the left and penalty games on the right. The data suggest that, across all games, simultaneous spillovers do occur. The high levels of cooperation by both Role C and Role A in penalty games (significantly greater than the *Penalty* treatment benchmark; $p-$ values < 0.001 for both first rounds and all rounds) appear to boost the cooperation rate of Role A when playing in no-penalty games. For both first rounds and all rounds Role A cooperates more on average than Role B, significant at the 10 and 1 percent levels, respectively. In turn, Role B is significantly more cooperative than the *Base game* benchmark rates, significant at the 1 percent level.

Looking at cooperation in the first repeated game, presented in the bottom panel, we clearly see evidence of learning. In the first round of the first game there is no significant difference in cooperation across treatments. Moreover, cooperation rates are all roughly 50%, suggesting that decisions are made randomly. When compared to behaviour across all repeated games, there is strong evidence of learn-

Table 6.2.4: PERCENTAGE OF COOPERATION IN TRIO GAMES

				ALL REPEATED GAMES						
	NO PENALTY					PENALTY				
	Base Game		Role B		Role A	Penalty		Role C		Role A
First Round	11.01	<***	20.13	<*	26.42	32.39	<***	49.69	>*	40.25
	∨***		∨***		∨***	∨***		∨***		∨***
All rounds	6.40	<***	13.75	<***	19.52	27.61	<***	35.53	>	33.65
				FIRST REPEATED GAME						
	NO PENALTY					PENALTY				
	Base Game		Role B		Role A	Penalty		Role C		Role A
First Round	55.56	>	41.18	<	43.75	55.56	<	56.25	<	41.18
	∨***		∨		∧	∨		∨		∨**
All rounds	26.67	<**	37.74	<	44.26	50.00	>	51.22	<	57.45

*, **, *** indicate significance at the 10, 5, and 1 percent level, respectively.

ing and a salient spillover effect emerging with experience.

Table 6.2.5 studies potential spillover effects under the *Trio* treatment, looking specifically at how the behaviour of either Role B or Role C might indirectly affect interactions between Role A and A's other partner.

Columns (1) is the baseline model, controlling the presence of a penalty in the match (i.e. controlling for interactions between A and C relative to those between A and B). In columns (2) - (4),the *Penalty in match* dummy is decomposed according to screen type, reflecting the unique legal environment experience by each role in a trio. In column (2), *Single screen with no penalty* is specified as the base level. Models (1) and (2) reveal a positive and significant effect of the presence of a penalty on the predicted probability, on average and for both screen types.

To isolate potential spillover effects, columns (3) and (4) examine how the effect of institutions (i.e. penalty and screen type) on cooperation to depend on the actions taking place in parallel by A's *other* opponent. To do so, columns (3) and (4) introduce the action taken in the first round of each match by Role A's other partner, and then restricts the sample to avoid endogeneity. Column (3) is thus, the predicted probability of cooperation for Role A and Role B (excluding C) controlling for Role C's choice in the first round. The interaction terms allow us to see

Table 6.2.5: Simultaneous Spillovers Probit Regressions

	Full sample		A and B	A and C
$P(\text{Cooperate}_i{=}1)$	(1)	(2)	(3)	(4)
Partner cooperated in round 1 of previous match	0.2241**	0.2373**	0.1597	0.2393
	(0.110)	(0.109)	(0.121)	(0.146)
Number of rounds in previous match	-0.0326	-0.0317	-0.0356	-0.0363
	(0.022)	(0.022)	(0.022)	(0.024)
Subject cooperated in round 1 of match 1	0.6308***	0.6355***	0.6592***	0.5533***
	(0.187)	(0.187)	(0.190)	(0.196)
Penalty in match	0.6126***			
	(0.099)			
Penalty screen§:				
Single screen with no penalty			-0.7700***	
			(0.254)	
Split screen with no penalty		0.2230	-0.3094**	
		(0.155)	(0.145)	
Split screen with penalty		0.7767***		0.6176***
		(0.136)		(0.126)
Single screen with penalty		0.6752***		0.5227***
		(0.136)		(0.151)
Role C cooperated in round 1			0.1730	
			(0.185)	
IT: *Single screen with no penalty*			0.0176	
× *Role C cooperated in round 1*			(0.365)	
IT: *Split screen with no penalty*			-0.4662*	
× *Role C cooperated in round 1*			(0.250)	
Role B cooperated in round 1				0.4280*
				(0.238)
IT: *Split screen with penalty*				-0.1537
× *Role B cooperated in round 1*				(0.241)
IT: *Single screen with penalty*				-0.2447
× *Role B cooperated in round 1*				(0.260)
Observations	2136	2136	1643	1566

Notes: § Base level specified as *Single screen no penalty* for column (2); *Split screen with penalty* for column (3); *Single screen with no penalty* for column (4).

Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

how Role C's choice depends on Penalty and Screen type. In other words, how the choice of C to cooperate in the first round of the match impacts, either directly or indirectly, the behaviour of the other players in their respective interactions. Column (4) is analogous, but for the effect of Role B on the behaviour of A and C.

The base levels for *Penalty screen* are specified as: *Split screen with penalty* for column (3); *Split screen with no penalty* for (4). As such, reported coefficients reflect the effect on cooperation relative to each base level, respectively. These specifications allow for direct comparisons across penalty type and within screen type.

In column (3) we see that both Role B and Role A in *Split screen with no penalty* are less likely to cooperate than Role A in *Split screen with penalty*, as we would expect. Interestingly, however, the magnitude of the difference is much greater for B than A. Moreover, the differential between the coefficients is significant at the 10 percent level. In the absence of any spillovers, we would expect this difference to zero. By simply changing the base level specification we evidence of potential spillover effects emerging.

Examining the indirect effect of C's behaviour on A and B, we find a significant negative effect on the interaction between *Split screen with no penalty* and Role C cooperating in round 1. Alone the effect does not reveal important information about the presence of spillovers since it is consistent with equilibrium profit-maximizing behaviour. Despite the lack of statistical significance, the positive coefficients on the interaction effects for *Split screen no penalty* is noteworthy. When we consider the two interaction terms together, it appears that C's actions have important indirect effects on the parallel interaction. The difference between the coefficients on the interaction terms, which should be zero if no spillovers exist, is large and significant at the 1 percent level.

Standard economic theory assumes that A's interactions with B and with C should be independent of one another. However, when we compare the effect of interacting in a no-penalty game against one opponent (i.e. Role B) to the effect of interacting in a no-penalty game on a split screen (i.e. simultaneously interacting in the penalty game), we see that Role A and Role B are indirectly impacted by Role C in very different ways. Simultaneous exposure to both different legal rules and to

the actions taken by each opponent cause Role A's behaviour to spillover across interactions. This suggests that when a subject is simultaneously exposed to distinct legal environments, experiences in one may influence decisions made in the other.

In column (4) Role B's behaviour in the first round has a positive and significant average effect on the predicted probability of cooperation for Role A and Role C. However, there is no significant difference between the conditional effect to Role B cooperating in round 1 for Role A and Role C. The indirect effect of B's behaviour on cooperation is less for both Role A and C that the direct effect on Role A in the no-penalty game. Though the interaction effects are not statistically significant, this result is consistent with the learning effects observed in the *Random - Duo* treatment. That is, the effect of B cooperating is more powerful in the no-penalty game where it sends a stronger prosocial signal than in the penalty game, where there is material incentive to cooperate.

6.3 Discussion

The results of Table 6.2.5 imply that spillovers tend to flow from penalty games to no-penalty games, and have a positive effect on the predicted probability of cooperation. One might argue that with more or less severe penalties and with different parameters of the game, the direction and magnitude of the spillovers might be altered. This underscores the complexity of attempting to isolate such inconspicuous effects, that are by nature the by-product of many other subtle influences on behaviour.

While much has been said about the potential crowding effects of sanctions, the literature on decision making in the context of formal incentives says little about behavioural spillovers from sequential and simultaneous learning under different legal rules, as explored in this paper. In the context of the legal pluralism of an increasingly complex and interconnected world, it is impossible to define a single cause of a particular spillover. However, future research should focus on identifying general factors and dynamics that tend to trigger positive or negative spillovers.

A better understanding of these features, would allow policymakers to design more efficient and effective legal incentives, and to eventually make use of complementary behavioural spillovers to particular laws aimed at overcoming collective action problems.

You must obey the law, always, not only when they grab you by your special place.

Vladimir Putin

7 Conclusion

The series of experiments presented in this paper sheds light on how formal institutions affect behaviour in an experimental setting. The evidence presented suggest that legal rules have a salient affect on motivating cooperation, and that these effects interact with the dynamics of different experiences. In particular, the explicit announcement of a penalty induces higher average levels of cooperation than when penalties are incorporated directly into the parameters of the game. This suggests that legal rules in our laboratory setting serve important expressive functions. By negatively framing the non-cooperative action, legal rules elicit various psychological reactions amongst subjects that result in levels of cooperation above those achieve by the deterrent function alone.

The results also suggest that these extra-deterrent effect of legal rules are not state-dependent. That is, past experiences in different legal environments spillover

into decisions made in interactions in the present. Conditional cooperation seems to be a crucial determinant in achieving high levels of cooperation in subsequent rounds. When combined with the additional framing effects of legal rules and a changing legal landscape, we see that subjects are initially very responsive to what happened in the immediate past, and that as experiences accumulate they become internalized as norm that tend to mitigate the volatility of behaviour.

When subjects interact in parallel across different legal environments, there seems to be an indirect effect of conditional cooperation from interactions in environments where legal rules exist to those in the absence of legal rules. However, additional work is required to more cleanly isolate these effects. Additional sessions well be needed to reduce confounding noise in the data. This is the next step in the research agenda. To isolate the information in the spillover treatments, we will need to control for the sequence of experience from penalty to no penalty games, and vice versa. We may also want to recruit a more homogeneous sample as an additional control. Why the heterogeneity of our sample is beneficial in terms of external validity, it adds unnecessary noise to the data.

Eventually, we hope to construct the foundation of a theoretical model of spillovers, which would significantly contribute to the growing literature.

In an era of increasing connectivity and global integration, in which time and space continue to condense, the question of spillovers is especially relevant.

8 Appendix A

8.1 Treatments parameters

T1: *Base game* treatment

$F = 10$ with probability $\alpha = 0$; $\delta = {}^{3}/_{4}$; $P(\text{Penalty in match}) = 0$. (C,C) is not supported in equilibrium.

T1 Payoff Matrix

		Player 2	
		C	D
Player 1	C	40 40	12 60
	D	12 60	35 35

T2: *Penalty* treatment

$F = 10$ with probability $a = 1$; $\delta = {}^{3}/_{4}$; $P(\text{Penalty in match}) = 1$.

(C,C) SPNE and risk dominant.

T2 Payoff Matrix

Player 2

Player 1	C	D
C	40 (P1), 40 (P2)	$60-aF$ (P1), 12 (P2)
D	$60-aF$ (P1), 12 (P2)	$35-aF$ (P1), $35-aF$ (P2)

T3: *Payoff equivalent* treatment

$F = 0$ with probability $a = 0$; $\delta = {}^{3}/_{4}$; $P(\text{Penalty in match}) = 0$.

(C,C) SPNE and risk dominant. Equivalent payoff function to T2; sanctions implicit in game parameters.

T3 Payoff Matrix

Player 2

Player 1	C	D
C	40 (P1), 40 (P2)	50 (P1), 12 (P2)
D	50 (P1), 12 (P2)	25 (P1), 25 (P2)

T4: *Random - Duo* treatment

$F = 10$ with probability $a = 1$; $\delta = {}^3/_4$; $P(\text{Penalty in match}) = 0.5$.

(C,C) not supported in equilibrium in T1 equivalent games. (C,C) SPNE and risk dominant in T2 equivalent games.

T5: *Trio* treatment

$F = 10$ with probability $a = 1$; $\delta = {}^3/_4$; $P(\text{Penalty in match}) = 0.5$.

(C,C) not supported in equilibrium in T1 equivalent games. (C,C) SPNE and risk dominant in T2 equivalent games.

$P(\text{Role A}) = P(\text{Role B}) = P(\text{Role C}) = {}^1/_3$.

Role A simultaneously plays T1 against B; T2 against C.

$\delta = {}^3/_4$.

T1 - A vs. B

Player 1 \ Player 2	C	D
C	40, 40	60, 12
D	60, 12	35, 35

T2 - A vs. C

Player 1 \ Player 2	C	D
C	40, 40	$60 - aF$, 12
D	$60 - aF$, 12	$35 - aF$, $35 - aF$

Table 8.1.1: Cooperation as SPNE and Risk Dominant Action

	No Penalty ($aF = 0$)	Penalty ($aF = 10$)
$\delta = {}^{3}/_{4}$	Neither SGPE nor RD	SGPE and RD
$\underline{\delta}$	0.8	0.4
δ^{*}	0.896	0.605

Table 8.1.1 shows the discount factors δ for which cooperation is an equilibrium and a risk dominant action in each treatment. As in Blonski and Spagnolo [2001], the critical values over which cooperation is supported in equilibrium and is a risk dominant action are denoted by $\underline{\delta}$ and δ^{*}, respectively.

8.2 Derivation of treatment parameters

Critical value of aF over which cooperation is not SPNE

To define the parameters of the benchmark *Base game* treatment where legal rules are absent, we reverse engineer the payoffs function in Dal Bó and Fréchette [2011] with $\delta = {}^{3}/_{4}$ and payoff to cooperation of 40. At these values cooperation is both payoff and risk dominant.

To do so, we calculate the minimum payoff to deviation that makes cooperation neither a risk dominant nor an equilibrium action, given $\delta = {}^{3}/_{4}$. The difference in these payoffs represents the minimum fine aF necessary to ensure that cooperation can be supported as a subgame-perfect equilibrium and risk dominant action.

$$\frac{40}{1-\delta} < (50 + aF) + \frac{(25 + aF)\delta}{1-\delta}$$

$$-25\delta - 10 < aF$$

Given a probability of continuation $\delta = {}^{3}/_{4}$, $aF > 8.75$. In other words, 8.75 is the

minimum value of the penalty to deviation that, when subtracted from the above payoffs, will yield stage game payoffs under which cooperation can no longer be supported in equilibrium. Given this result, we define $aF = 10$, where $F = 10$ with probability $a = 1$, such that deviation is strictly dominated when $\delta = {}^{3}/_{4}$.

Critical value of aF over which defection is risk dominant

According to the simplified interpretation of risk dominance of Blonski and Spagnolo [2001], "cooperation is risk dominant if playing 'grim trigger' (G) is the best response to other player's choosing G or AD ('always defect') with equal probabilities" (Dal Bó and Fréchette, 2011, 416).

To reverse engineer our baseline PD stage game from Dal Bó and Fréchette [2011], we determine the minimal fine necessary for which deviation is risk dominant. Let $u_i(\phi)$ represent the set of all cooperation equilibrium and $v_i(\phi)$ represent the set of all deviation equilibrium. For deviation to be risk dominant, the 'Nash-product' [Harsanyi and Selten, 1988] of the cooperation equilibria $\Big(u_1(\phi)u_2(\phi)\Big)$ must strictly dominated by the 'Nash-product' of the deviation equilibria $\Big(v_1(\phi)v_2(\phi)\Big)$.

$$u_i(\phi)u_2(\phi) < v_1(\phi)v_2(\phi)$$

$$\begin{aligned}
u_i(\phi) :& \frac{40}{1-\delta} - (50 + aF) - \frac{\delta(25 + aF)}{1-\delta} \\
v_i(\phi) :& \frac{(25 + aF)}{1-\delta} - 12 - \frac{\delta(25 + aF)}{1-\delta} = (25 + aF) - 12
\end{aligned}$$

$$\left[\frac{40}{1-\delta} - (50 + aF) - \frac{\delta(25 + aF)}{1-\delta}\right]^2 < \left[(25 + aF) - 12\right]^2$$

$$\left[\frac{25\delta - 10 - aF}{1-\delta}\right]^2 < (13 + aF)^2$$

$$\frac{23 - 38\delta}{\delta - 2} < F\delta$$

Therefore, given $\delta = {}^3/_4$, $aF > 4.4$. In other words, for deviation to be risk dominant, the payoff to D must be at least 4.4 monetary units greater than the stage game payoffs of Dal Bó and Fréchette [2011] where the payoff to cooperation is 40 and $\delta = {}^3/_4$. Thus, the fine $aF = 10$ does indeed make defection the payoff and risk dominant action in the *Base game* treatment.

9
Appendix B

9.1 Probit regression control variables

Table 9.1.1 presents the baseline independent variables of the probit regression analysis of predicted probability of cooperation. They include four individual player characteristics controls and three game-play dynamics argued to be relevant to cooperation Dal Bó and Fréchette [2011].

Table 9.1.1: Baseline Probit Regression Independent Variables

Variable	Definition	Sample Average
Participant Characteristics		
Gender	1 if subject is male, 0 if female	0.55
Student	1 if subject is a student, 0 otherwise	0.48
Age	Subject's age	33.14
Risk aversion	Scale of preference from 0 (avoid risk) to 10 (take risk)	5.25
Game Characteristics		
Partner cooperated in round 1 of previous match	1 if partner cooperated in first round of previous match, 0 otherwise	0.26
Number of rounds in previous match	Total number of rounds in previous match	3.50
Subject cooperated in round 1 of match 1	1 if subject cooperated in first round of the first match, 0 otherwise	0.56

Basin of attraction

When deciding whether it is optimal to pursue a cooperative strategy, such as *grim trigger* (G), or a defecting strategy, such as *always defect* (AD), a player considers both the fundamental parameters of the game and his or her beliefs about the expected probability that his or her opponent will cooperate or play AD[1]. In Figure 6.1.1, the horizontal dotted line represents the belief of the partner's propensity to play the cooperative strategy that leaves the subject indifferent between the cooperative and defecting strategies in penalty games. In no-penalty games, it is never optimal to cooperate.

Given $\delta = {}^{3}/_{4}$ and the payoff function of the Penalty and Payoff equivalent games, players are indifferent between strategies if they believe their opponent will choose the cooperative strategy with probability 0.27. This critical value denotes the limit of the basin of attraction of AD (the set of beliefs that make AD optimal).Given $\delta = {}^{3}/_{4}$ and the payoff function of the penalty games, players are indifferent between strategies if they believe their opponent will choose the coop-

[1]G is never a best response for No Penalty treatments.

erative strategy with probability 0.27. This critical value denotes the limit of the basin of attraction of AD (the set of beliefs that make AD optimal).

Proof. Suppose player *i* believes his or her opponent will play G with probability *p* and AD with probability $(1-p)$, and that the play attempts to cooperate using *grim trigger* strategy:

I. Play *C* in every period unless someone plays *D*, in which go to **II**.

II. Play *D* forever.

Suppose up to time *t*, *D* has never been played. Then *i*'s expected payoffs are:

$$\begin{aligned}
\text{Play G} \quad &\Rightarrow \quad p\left(\frac{40}{1-\delta}\right) + (1-p)\left[12 + \frac{25\delta}{1-\delta}\right] \\
\text{Play AD} \quad &\Rightarrow \quad p\left[50 + \frac{25\delta}{1-\delta}\right] + (1-p)\left(\frac{25}{1-\delta}\right)
\end{aligned}$$

If $\delta = {}^{3}/_{4}$, the distribution *p* of cooperative types that leaves player *i a priori* indifferent between strategies is given by,

$$\begin{aligned}
p\left(\frac{40}{1-\delta}\right) + (1-p)\left[12 + \frac{25\delta}{1-\delta}\right] &= p\left[50 + \frac{25\delta}{1-\delta}\right] + (1-p)\left(\frac{25}{1-\delta}\right) \\
p(160) + (1-p)(87) &= p(125) + (1-p)(100) \\
87 + 73p &= 100 + 25p \\
p &= \frac{13}{48} \\
p &\approx 0.27
\end{aligned}$$

so if a player believes his or her opponent will choose a cooperative strategy with probability greater than or equal to 0.27, it is optimal to also play G.

9.2 Session instructions

Bellow are the relevant game instructions provided at the beginning session, and the and questionnaire that immediately followed the experiment.

***Base game* treatment instructions**

BIENVENUE

Avant de commencer, merci d’éteindre vos téléphones portables.

Vous allez participer à une expérience sur la prise de décision. Vous serez payés immédiatement à la fin de l’expérience. Merci de rester assis jusqu’à ce qu’on appelle votre numéro de poste.

Vos gains à la fin de l’expérience dépendront en partie de vos choix, en partie du choix des autres participants et en partie de tirages aléatoires.

L'expérience entière aura lieu par le biais de terminaux informatiques, et toutes les interactions entre vous aura lieu par les ordinateurs. Merci de ne pas essayer de communiquer avec les autres participants pour toute la durée de l’expérience.

INSTRUCTIONS

Nous allons maintenant lire les instructions. Si vous avez une quelconque question, levez votre main et nous répondrons à la question de telle sorte que tout le monde puisse entendre.

INSTRUCTIONS GENERALES

A) L’ENVIRONEMENT GENERAL

1. Dans cette expérience vous allez devoir prendre des décisions en plusieurs séries de tours. Nous allons appeler chaque série de tours un jeu. Au début de chaque jeu un partenaire vous est aléatoirement attribué. Chaque tour à l’intérieur d’un même jeu est joué avec cette même personne. En revanche, quand un jeu se termine, vous débuterez un nouveau jeu **avec un autre participant.**

2. La durée d’un jeu est déterminée aléatoirement de la façon suivante. A la fin de chaque tour, un tirage est effectué. Dans ce tirage, il y a 75% de chances que le jeu continue pour un tour supplémentaire. Par exemple, si vous êtes au tour 2, la probabilité d’avoir un troisième tour dans ce jeu est de 75%. Si vous êtes au tour 3, la probabilité d’avoir un quatrième tour est de 75% et ainsi de suite. **Cette probabilité est indépendante de vos choix ou de ceux des autres. Chaque jeu peut être très court ou très long : ceci est totalement aléatoire et ne dépend pas de vos choix.**

   Si le tirage est un succès et donne lieu à un tour supplémentaire, ce symbole s'affichera sur l'écran : 

   Si le tirage est un échec et le jeu se termine, ce symbole s'affichera sur l'écran : 

3. Quand un jeu se termine, un nouveau jeu commence avec un nouveau participant et ce nouveau jeu se déroule de la même manière.

4. Durée de l’expérience: vous ne commencerez pas un nouveau jeu après la 15ème minute mais vous finirez le jeu en cours.

5. UNE FOIS QUE L’EXPÉRIENCE EST TERMINÉE POUR VOUS, MERCI DE RESTER ASSIS ET DE NE PAS PARLER !

B) CHOIX ET GAINS DANS CHAQUE TOUR

Chaque tour, chaque joueur devra choisir entre deux options X et Y. Ce choix est simultané : le joueur n’observe pas le choix de l’autre avant de choisir. Le choix des deux joueurs détermine leurs gains dans ce tour de la façon suivante :

- Si les deux joueurs choisissent **X**, les deux obtiennent 40 jetons.
- Si vous choisissez **X** et l'autre **Y**, vous obtenez 12 et l'autre obtient 60 jetons.
- Si vous choisissez **Y** et l'autre **X**, vous obtenez 60 et l'autre obtient 12 jetons.
- Si les deux joueurs choisissent **Y**, les deux obtiennent 35 jetons.

Les gains peuvent être résumés dans la table suivante où la première entrée de chaque cellule représente votre gain, tandis que la deuxième entrée représente le gain de votre partenaire.

	CHOIX DU PARTENAIRE	
	X	**Y**
VOTRE CHOIX **X**	40, 40	12, 60
VOTRE CHOIX **Y**	60, 12	35, 35

A la fin de chaque tour un écran affiche tous les choix ainsi que les gains de chacun. A la fois les joueurs connaîtront le choix de l’autre.

Rappel: une fois que les gains dans ce tour auront été montrés, un tirage aléatoire déterminera si un autre tour dans ce jeu sera effectué (comme nous l’avons expliqué dans la partie A). Après chaque tour il y a 75% de chances d’avoir un autre tour dans ce jeu.

C) PAIEMENT

A la fin de l'expérience, vous serez payé 1 euro tout le 100 point gagné durant l'expérience.

D) QUESTIONS

Avez-vous des questions?

Avant de débuter, nous vous rappelons que:

- La durée du jeu est déterminée aléatoirement.
- Après chaque tour il y a 75% de chances d'avoir un autre tour dans le même jeu.
- Vous jouerez avec la même personne tout au long de chaque jeu.
- Quand un jeu se termine, un nouveau jeu commence avec un nouveau participant et ce nouveau jeu se déroule de la même manière.
- Les différents jeux sont indépendants. En particulier les choix effectués dans un jeu n'affectent pas les autres jeux.

***Penalty* treatment instructions**

B) CHOIX ET GAINS DANS CHAQUE TOUR

Chaque tour, chaque joueur devra choisir entre deux options X et Y. Ce choix est simultané : le joueur n’observe pas le choix de l’autre avant de choisir. Le choix des deux joueurs détermine leurs gains dans ce tour de la façon suivante :

- Si les deux joueurs choisissent **X**, les deux obtiennent 40 jetons.
- Si vous choisissez **X** et l'autre **Y**, vous obtenez 12 et l'autre obtient 60 jetons.
- Si vous choisissez **Y** et l'autre **X**, vous obtenez 60 et l'autre obtient 12 jetons.
- Si les deux joueurs choisissent **Y**, les deux obtiennent 35 jetons.

Les gains peuvent être résumés dans la table suivante où la première entrée de chaque cellule représente votre gain, tandis que la deuxième entrée représente le gain de votre partenaire.

		CHOIX DU PARTENAIRE	
		X	**Y**
VOTRE CHOIX	**X**	40, 40	12, 60
	Y	60, 12	35, 35

En plus de ces gains, si un joueur joue Y, il sera pénalisé de 10 jetons.

A la fin de chaque tour un écran affiche tous les choix ainsi que les gains de chacun (y compris si l’autre joueur a reçu une pénalité)

***Payoff equivalent* treatment instructions**

B) CHOIX ET GAINS DANS CHAQUE TOUR

Chaque tour, chaque joueur devra choisir entre deux options X et Y. Ce choix est simultané : le joueur n'observe pas le choix de l'autre avant de choisir. Le choix des deux joueurs détermine leurs gains dans ce tour de la façon suivante :

- Si les deux joueurs choisissent **X**, les deux obtiennent 40 jetons.
- Si vous choisissez **X** et l'autre **Y**, vous obtenez 12 et l'autre obtient 50 jetons.
- Si vous choisissez **Y** et l'autre **X**, vous obtenez 50 et l'autre obtient 12 jetons.
- Si les deux joueurs choisissent **Y**, les deux obtiennent 25 jetons.

Les gains peuvent être résumés dans la table suivante où la première entrée de chaque cellule représente votre gain, tandis que la deuxième entrée représente le gain de votre partenaire.

		CHOIX DU PARTENAIRE	
		X	**Y**
VOTRE CHOIX	**X**	40, 40	12, 50
	Y	50, 12	25, 25

A la fin de chaque tour un écran affiche tous les choix ainsi que les gains de chacun.

***Random - Duo* treatment instructions**

B) CHOIX ET GAINS DANS CHAQUE TOUR

Chaque jeu peut être de deux types: jeu **Sans Pénalité** ou jeu **Avec Pénalité**. Au début de chaque jeu, le type sera déterminé aléatoirement : le jeu sera de **Type Sans Pénalité** avec 1 chance sur 2 et de **Type Avec Pénalité** avec 1 chance sur 2. Le Type du jeu restera le même pour chaque tour dans ce jeu.

Quel que soit le type de jeu, à chaque tour, chaque joueur devra choisir entre deux options X et Y. Ce choix est simultané : le joueur n'observe pas le choix de l'autre avant de choisir. Le choix des deux joueurs détermine leurs gains dans ce tour de la façon suivante :

- Si les deux joueurs choisissent **X**, les deux obtiennent 40 jetons.
- Si vous choisissez **X** et l'autre **Y**, vous obtenez 12 et l'autre obtient 60 jetons.
- Si vous choisissez **Y** et l'autre **X**, vous obtenez 60 et l'autre obtient 12 jetons.
- Si les deux joueurs choisissent **Y**, les deux obtiennent 35 jetons.

Les gains peuvent être résumés dans la table suivante où la première entrée de chaque cellule représente votre gain, tandis que la deuxième entrée représente le gain de votre partenaire.

VOTRE CHOIX \ CHOIX DU PARTENAIRE	**X**	**Y**
X	40, 40	12, 60
Y	60, 12	35, 35

La seule différence entre le jeu de type avec pénalité et le type sans pénalité est la suivante. En plus de ces gains, si le type du jeu est « Type Avec Pénalité », si un joueur joue Y, il sera pénalisé de 10 jetons.

A la fin de chaque tour un écran affiche tous les choix ainsi que les gains de chacun (y compris si l'autre joueur a reçu une pénalité)

***Trio* treatment instructions**

INSTRUCTIONS GENERALES

A) L'ENVIRONEMENT GENERAL

1. Dans cette expérience vous allez devoir prendre des décisions en plusieurs séries de tours. Nous allons appeler chaque série de tours un jeu.

2. Au début de chaque jeu vous serez aléatoirement organisés en groupe de 3 joueurs. Chaque tour à l'intérieur d'un même jeu est joué entre le même groupe de participants. En revanche, quand un jeu se termine, vous débuterez un nouveau jeu **avec un autre groupe.**

3. Au début de chaque jeu, un rôle (rôle A, rôle B ou rôle C) vous est aléatoirement attribué. Chaque rôle est équiprobable. Vous avez une chance sur trois d'être le rôle A, une chance sur trois d'être le rôle B et une chance sur trois d'être le rôle C. Les choix et gains de chaque rôle sont définis ci-dessous. Chaque tour, le rôle A aura deux choix à faire, un affectant le rôle B et un affectant le rôle C. Les rôles B et C auront un seul choix à faire et ces choix affecteront uniquement le rôle A. Ces choix seront détaillés ci-dessous dans la section choix et gains dans chaque tour.

4. La durée d'un jeu est déterminée aléatoirement de la façon suivante. A la fin de chaque tour, un tirage est effectué. Dans ce tirage, il y a 75% de chances que le jeu continue pour un tour supplémentaire. Par exemple, si vous êtes au tour 2, la probabilité d'avoir un troisième tour dans ce jeu est de 75%. Si vous êtes au tour 3, la probabilité d'avoir un quatrième tour est de 75% et ainsi de suite. **Cette probabilité est indépendante de vos choix ou de ceux des autres. Chaque jeu peut être très court ou très long : ceci est totalement aléatoire et ne dépend pas de vos choix.** Cette procédure est faite indépendamment pour le jeu entre A et B et pour le jeu entre A et C. Le nombre de tour dans le jeu entre A et B pourra donc être différent du nombre de tour dans le jeu entre A et C.

   Si le tirage est un succès et donne lieu à un tour supplémentaire, ce symbole s'affichera sur l'écran :

   

   Si le tirage est un échec et le jeu se termine, ce symbole s'affichera sur l'écran :

   

5. Quand un jeu se termine, un nouveau jeu commence avec un nouveau groupe et ce nouveau jeu se déroule de la même manière.

6. Durée de l'expérience: vous ne commencerez pas un nouveau jeu après la 15ème minute mais vous finirez le jeu en cours.

7. UNE FOIS QUE L'EXPÉRIENCE EST TERMINÉE POUR VOUS, MERCI DE RESTER ASSIS ET DE NE PAS PARLER !

B) DEBUT DU JEU

On rappelle qu'au début de chaque jeu, un rôle (rôle A, rôle B ou rôle C) vous est aléatoirement attribué. Chaque rôle est équiprobable. Vous avez une chance sur trois d'être le rôle A, une chance sur trois d'être le rôle B et une chance sur trois d'être le rôle C. Les choix et gains de chaque rôle sont définis ci-dessous

Tous les jeux sont indépendants. Le rôle attribué dans ce jeu et les choix faits dans ce jeu n'auront aucune influence sur les jeux futurs.

C) CHOIX ET GAINS DANS CHAQUE TOUR

Dans chaque tour, le rôle A aura deux choix à faire, un affectant le rôle B et un affectant le rôle C. Les rôles B et C auront un seul choix à faire et ces choix affecteront uniquement le rôle A. Tous les choix dans un tour sont simultanés: chaque joueur n'observe pas le choix des autres avant de choisir.

Les rôles B et C choisissent entre **X** et **Y**
Le rôle A a deux choix à faire, entre **X** et **Y pour son interaction avec B** d'une part et **X** et **Y pour son interaction avec C** d'autre part. Son écran sera divisé en deux parties, l'un correspondant à son interaction avec le rôle B (sur fond bleu), l'autre avec C (sur fond vert).

Les gains dans l'interaction entre le rôle A et le rôle B dans chaque tour est défini de la façon suivante. Ces gains ne dépendent pas du choix du rôle C ou du choix du rôle A dans son interaction avec C.

- Si le rôle A et le rôle B choisissent X les deux obtiennent 40 jetons.
- Si le rôle A choisit **X** et le rôle B **choisit Y**, le rôle A obtient 12 et le rôle B obtient 60 jetons.
- Si le rôle A choisit **Y** et le rôle B **choisit X,** le rôle A obtient 60 et le rôle B obtient 12 jetons.
- Si le rôle A et le rôle B choisissent Y les deux obtiennent 35 jetons.

Les gains peuvent être résumés dans la table suivante où la première entrée de chaque cellule représente le gain du rôle A, tandis que la deuxième entrée représente le gain du rôle B.

CHOIX DE **B**

CHOIX DE **A**	**X**	**Y**
X	40, 40	12, 60
Y	60, 12	35, 35

Dans l'interaction entre le rôle A et le rôle C, les gains sont définis de la même manière, selon la table suivante. Les gains ne dépendent pas du choix du rôle B ou des choix du rôle A dans son interaction avec B :

CHOIX DE **C**

CHOIX DE **A**	**X**	**Y**
X	40, 40	12, 60
Y	60, 12	35, 35

Mais en plus de ces gains, si un joueur joue Y, il sera pénalisé de 10 jetons.

Sur l'écran au début d'un jeu :

- Si vous êtes dans le rôle B, on vous indiquera **« JEU SANS PENALITE »**
- Si vous êtes dans le rôle C, on vous indiquera **« JEU AVEC PENALITE »**
- Si vous êtes dans le rôle A, votre écran sera coupé en deux. Sur fond bleu vous jouez avec B et il sera indiqué **« JEU SANS PENALITE ».** Sur fond vert vous jouez avec C et il sera indiqué **« JEU AVEC PENALITE »**

Questionnaire

1. Sexe : ☐ H ☐ F
2. Âge : ________
3. Étudiant : ☐ Oui
4. Diplôme (préparé) : ________________________________
5. Emploi (poste occupé) : ________________________________

Parmi les choix suivants, quelle était le facteur déterminant pour votre choix d'investissement au premier tour quand vous aviez été attribué le rôle A ?

6. Choix attendu dans les tours suivants du rôle B
   - ☐ moins important
   - ☐ moyennement important
   - ☐ important
   - ☐ plus important
7. Montant du prix espéré
   - ☐ moins important
   - ☐ moyennement important
   - ☐ important
   - ☐ plus important
8. Probabilité d'avoir beaucoup de tours dans ce jeu
   - ☐ moins important
   - ☐ moyennement important
   - ☐ important
   - ☐ plus important
9. Autre (spécifiez) : ________________________________
   - ☐ moins important
   - ☐ moyennement important
   - ☐ important
   - ☐ plus important

Parmi les choix suivants, quelle était les facteurs déterminants pour votre choix entre Haut et Bas (après un succès dans la loterie attribuant le prix) quand vous aviez le rôle B ?

10. Montant du prix espéré
    - ☐ moins important
    - ☐ moyennement important
    - ☐ important
    - ☐ plus important
11. Réaction attendue du rôle A
    - ☐ moins important
    - ☐ moyennement important
    - ☐ important
    - ☐ plus important

12. Probabilité d'avoir beaucoup de tours dans ce jeu
    - ☐ moins important
    - ☐ moyennement important
    - ☐ important
    - ☐ plus important

13. Autre (spécifiez) : ______________________________
    - ☐ moins important
    - ☐ moyennement important
    - ☐ important
    - ☐ plus important

Comment vous décririez-vous ? Êtes-vous quelqu'un qui en général est disposé à prendre des risques ou plutôt quelqu'un qui évite le risque ?

Choisissez votre réponse sur cette échelle de 0 (évite le risque) à 10 (prends des risques).

14. évite le risque ☐—☐—☐—☐—☐—☐—☐—☐—☐—☐ prends des risques

References

J. Andreoni. An experimental test of the public goods crowding-out hypothesis. *American Economic Review*, 83(5):1317–1327, 1993.

J. Andreoni. Cooperation among strangers under the shadow of the future. *American Economic Review*, 99(3):979–1005, 2009.

James Andreoni and Hal Varian. Preplay contracting in the prisoners' dilemma. *Economic Sciences*, 96:10933–10938, 1999.

Robert Axelrod. An evolutionary approach to norms. *The American Political Science Review*, 80(4):1095–1111, 1986.

Michael Bacharach and Michele Bernasconi. The variable frame theory of focal points: An experimental study. *Games and Economic Behavior*, 19(1):1–45, 1997.

Jean-Pierre Benoit, Roberto Galbiati, and Emeric Henry. Rational Parasites. CEPR Discussion Paper No. DP9351, February 2013.

Maria Bigoni, Jan Potters, and Giancarlo Spagnolo. Flexibility and collusion with imperfect monitoring. CEPR Discussion Paper No. 8877, 2012.

Matthias Blonski and Giancarlo Spagnolo. Prisoners' other dilemma. Working Paper Series in Economics and Finance 437, Stockholm School of Economics, February 2001.

Matthias Blonski, Peter Ockenfels, and Giancarlo Spagnolo. Co-operation in infinitely repeated games: Extending theory and experimental evidence, 2007.

I. Bohnet and B. Cooter. Expressive law: Framing or equilibrium selection? Working Paper No. RWP03-046, Harvard University, John F. Kennedy School of Government, 2003.

Iris Bohnet and Yael Baytelman. Institutions and trust: Implications for preferences, beliefs and behavior. *Rationality and Society*, 19(1):99–135, 2007.

Iris Bohnet, Bruno S. Frey, and Steffen Huck. More order with less law: On contract enforcement, trust and crowding. *American Political Science Review*, 95(1): 131–144, 2001.

Iris Bohnet, Fiona Greig, Benedikt Herrmann, and Richard Zeckhauser. Betrayal aversion: Evidence from brazil, china, oman, switzerland, turkey, and the united states. *American Economic Review*, 98(1):294–310, 2008.

Pedro Bordalo, Nicola Gennaioli, and Andrei Shleifer. Salience theory of choice under risk. *The Quarterly Journal of Economics*, 127(3):1243–1285, 2012.

Samuel Bowles and Sandra Polaní-Reyes. Economic incentives and social preferences: Substitutes or complements? *Journal of Economic Literature*, 50(2): 368–425, 2012.

Lisa Bruttel and Tim Friehe. On the path-dependence of tax compliance. Research Paper Series 59, Thurgau Institute of Economics and Department of Economics at the University of Konstanz, 2010.

Juan-Camilo Cardenas. Norms from outside and from inside: an experimental analysis on the governance of local ecosystems. *Forest Policy and Economics*, 6 (3–4):229–241, 2004.

Juan Camilo Cardenas, John Stranlund, and Cleve Willis. Local environmental control and institutional crowding-out. *World Development*, 28(10):1719–1733, 2000.

André Casajus. Focal points in framed strategic forms. *Games and Economic Behavior*, 32(2):263–291, 2000.

Andy Clark and Annette Karmiloff-Smith. The cognizer's innards: A psychological and philosophical perspective on the development of thought. *Mind and Language*, 8(4):487–519, 1994.

Andrew M. Colman. Salience and focusing in pure coordination games. *Journal of Economic Methodology*, 4(1):61–81, 1997.

Robert Cooter. Expressive law and economics. *The Journal of Legal Studies*, 27 (S2):585–607, 1998.

Pedro Dal Bó. Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games. *American Economic Review*, 95(5):1591–1604, December 2005.

Pedro Dal Bó and Guillaume R. Fréchette. The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Economic Review*, 101(1):411–429, 2011.

John Duffy and Jack Ochs. Cooperative behavior and the frequency of social interaction. *Games and Economic Behavior*, 66(2):785–812, July 2009.

J. Falkinger, Ernst Fehr, Simon Gächter, and R. Winter-Ebmer. A simple mechanism for the efficient private provision of public goods - experimental evidence. *American Economic Review*, 90:247–264, 2000.

Ernst Fehr and Simon Gächter. Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994, 2000.

Ernst Fehr and Simon Gächter. Do incentive contracts crowd out voluntary cooperation? IEW - Working Papers 034, Institute for Empirical Research in Economics - University of Zurich, 2002.

Robert M. Feinberg and Thomas A. Husted. An experimental test of discount-rate effects on collusive behaviour in duopoly markets. *The Journal of Industrial Economics*, 41(2):153–160, 1993.

Raymond Fisman and Edward Miguel. Cultures of corruption: Evidence from diplomatic parking tickets. NBER Working Papers 12312, National Bureau of Economic Research, Inc, June 2006.

Bruno S. Frey and Reto Jegen. Motivation crowding theory: A survey of empirical evidence. *Journal of Economic Surveys*, 15(5):589–611, 2001.

Drew Fudenberg and Eric Maskin. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, 54(3):533–54, May 1986.

Drew Fudenberg and Jean Tirole. *Game Theory*, volume 1 of *MIT Press Books*. The MIT Press, July 1991.

Drew Fudenberg, David G. Rand, and Anna Dreber. Slow to anger and fast to forgive: Cooperation in an uncertain world. *American Economic Review*, 102 (2):720–749, 2012.

P. Funk. Is there an expressive function of law? an empirical analysis of voting laws with symbolic fines. *American Law and Economic Review*, 9(1):135–159, 2007.

Gerd Gigerenzer and Wolfgang Gaissmaier. Heuristic decision making. *Annual Review of Psychology*, 62:451–482, 2011.

Uri Gneezy and Aldo Rustichini. A fine is a price. *Journal of Legal Studies*, 29(1): 1–17, 2000.

John C. Harsanyi and Reinhard Selten. *A General Theory of Equilibrium Selection in Games*, volume 1. The MIT Press, December 1988.

Douglas D. Heckathorn. Collective sanctions and the creation of prisoner's dilemma norms. *American Journal of Sociology*, 94(3):535–562, 1988.

D. M. Kahan. Social influence, social meaning and deterrence. *Virginia Law Review*, 3:349–395, 1997.

Daniel Kahneman. Maps of bounded rationality: Psychology for behavioral economics. *The American Economic Review*, 93(5):1449–1475, 2003.

Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292, 1979.

R.M. Krauss, J.L. Freedman, and M. Whitcup. Field and laboratory studies of littering. *Journal of Experimental Social Psychology*, 14(1):109–122, 1978.

George J. Mailath and Larry Samuelson. *Repeated Games and Reputations: Long-Run Relationships*. OUP USA. Oxford University Press, 2006.

Richard H. McAdams. A focal point theory of expressive law. *Virginia Law Review*, 86:1649–1729, 2000.

Richard H. McAdams. Beyond the prisoners' dilemma: Coordination, game theory, and law. *Southern California Law Review*, 82, 2009. U of Chicago Law & Economics, Olin Working Paper No. 437.

Richard H. McAdams and Janice Nadler. 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(1):87–123, 2005.

J. Keith Murnighan and Alvin E. Roth. Expecting continued play in prisoner's dilemma games: A test of several models. *The Journal of Conflict Resolution*, 27 (2):279–300, 1983.

Mancur Olson. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press, Cambridge, MA, 1965.

E. Ostrom. *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press, Cambridge, 1990.

Elinor Ostrom. Collective action and the evolution of social norms. *The Journal of Economic Perspectives*, 14(3):137–158, 2000.

Thomas R. Palfrey and Howard Rosenthal. Repeated play, cooperation and coordination: An experimental study. *Review of Economic Studies*, 61(3):545–565, 1994.

Alexander Peysakhovich and David G. Rand. Habits of virtue: Creating norms of cooperation and defection in the laboratory. Working papers series, October 2013.

David G. Rand and Martin A. Nowak. Human cooperation. *Trends in Cognitive Sciences*, 17(8):413–425, 2013.

Alvin E. Roth and J.Keith Murnighan. Equilibrium behavior and repeated play of the prisoner's dilemma. *Journal of Mathematical Psychology*, 17(2):189–198, 1978.

Ido Shahar. A Tale of Two Courts: How Organizational Ethnography Can Shed New Light on Legal Pluralism. *PoLAR: Political and Legal Anthropology Review*, 36(1):118–137, 2013.

A. W. Tucker. On jargon: The prisoner's dilemma. *UMAP Journal*, 1:101, 1980.

J. R. Tyran and L. Feld. Achieving compliance when legal sanctions are non-deterrent. *Scandinavian Journal of Economics*, 108(1):135–156, 2006.

Chen-Bo Zhong, Jeffrey Loewenstein, and J. Keith Murnighan. Speaking the same language: The cooperative effects of labeling in the prisoner's dilemma. *The Journal of Conflict Resolution*, 51(3):431–456, 2007.

Colophon

This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The body text is set in 11 point Arno Pro, designed by Robert Slimbach in the style of book types from the Aldine Press in Venice, and issued by Adobe in 2007. A template, which can be used to format a PhD thesis with this look and feel, has been released under the permissive MIT (X11) license, and can be found online at github.com/suchow/ or from the author at suchow@post.harvard.edu.