Complementary Institutions and Economic Development: An Experimental Study

Andrew Kloosterman and Andrew Schotter

July 14, 2016

Abstract

This paper considers the problem of why societies develop differently, a question most recently articulated by Acemoglu and Robinson (2012). We follow North (1990) in defining institutions as the "rules of the game in society." The question then becomes why do some societies develop functional institutions while others do not? To investigate this question, we develop and examine a specific type of dynamic game (which we call an Institutional Game). Our point is that complementarities among the choices that all societies make as they develop can help to answer this question.

Key words: Economic Development, Dynamic Games, Institutions

JEL Codes: C72, C91, O12

The authors would like to thank Alberto Bisin, Jon Eguia, and participants of the NYU CESS Informal Lunch Seminar, Berlin Behavioral Economics Seminar, the June 2012 ESA International Meetings session on Game Theory, and May 2015 3rd NYU Alumni Conference. We would also like to thank the editors of Games and Economic Behavior and the anonymous referees for their faith in this project and their helpful suggestions.

"Institutions are the rules of the game in a society or, more formally, are the humanly devised constraints that shape interaction" (North 1990, p.3)

1 Introduction

Ever since Adam Smith, one perennial problem for economists is the question of why some economies develop and others do not. Acemoglu, Johnson and Robinson (2005) and Acemoglu and Robinson (2012) list three possible explanations. As they explain, the "geography school" would claim that some economies are blessed with better climates and geographical conditions which makes agriculture and engaging in economic activities easier. A second view is that differences in development are the result of differences in the culture of the agents and the norms and conventions they develop to govern their work and social life (see Weber (1930), Greif (1994), Fernandez (2008, 2012)). Finally, there is the institutional view, of which Acemoglu and Robinson (2012), along with North (1990), North and Thomas (1973), Williamson (1985) and others are the main advocates, which claims that differences in economic development can be best attributed to differences in the institutions that different societies develop. For Acemoglu and Robinson, those societies that develop or inherit efficient economic institutions, specifically those that foster strong property rights and proper incentives, prosper while those whose institutions are dysfunctional suffer.

We view economic (and political) institutions as the set of rules (game forms) by which economic (and political) activity is governed. As quoted above, North (1990) states that economic institutions are "the rules of the game in a society or, more formally, are the humanly devised constraints that shape interaction." Using this as our definition, the question then becomes why are some countries playing different games than others and why are some playing good games (i.e. games that lead to efficient outcomes) and others playing bad or dysfunctional games (games whose strategy sets are limited or only contain inferior equilibria)?

For Acemoglu and Robinson (2012), institutions are the end product of social conflict where different groups in society pursue their own selfish interests sometimes at the expense of the greater good. When elites are in power (or where political power is not shared broadly), the groups in control have the ability to seek rents at the expense of the general population and this prevents the proper set of property rights and incentives from being established. Sometimes these elites are the remnants of previous colonial experiences and sometimes they emerge on their own, but the set of institutions existing today is an outcome of this dynamic political economy exercise.

We agree that institutions often arise in this way, but take a more general approach following Schotter (1981) who looks at the emergence of social institutions. In that analysis, the set of institutions we have today is the result of decisions made in the past to solve a set of problems that all societies face. The solutions to these common problems determine the institutional structure of the economy today, an object we call the "Final Game." For example, all societies must, at some point as they evolve, decide on a legal system (common or civil), (see Glaeser and Shliefer (2002), LaPorta et al. (1998)), a banking system, a religion (Judaism, Christianity, Islam, Hinduism, etc.), a political system (monarchy, parliamentary democracy, etc.), a set of inheritance laws (primogeniture or ultimogeniture), a system of property rights (common or private), a constitution, family structure (liberal or authoratarian) (see Todd (1983, 1990)) (strong or weak) (Alesina and Giuliano (2010, 2015)), etc. We will call these problems, which in our analysis are themselves games, the "Formative Games." In total, the Formative Games and Final Game constitute a dynamic game we call the Institutional Game.

The interesting point, and focus of this paper, is that the solutions chosen to solve one of these problems places constraints on the strategies the society has when facing other problems. For example, say that a society decides to adhere to Islam in the "Formative Game" where religion is chosen. This solution, however, has implications for the type of banking system they can then employ since, under Islam, interest is not allowed and a whole set of rules become relevant when conducting Islamic banking (see Kuran (2011)).[1]

A further example may be the choice of legal systems which, as LaPorta et al. (1998) have demonstrated, may have significant consequences for the way debtors and lenders and property rights are protected. Here, how a society decides on the legal system places constraints on the way stockholders are protected under bankruptcy and hence alters the strategies available to agents in other financial games they play. In other words, the existence of a solution to one societal formative problem creates constraints on the strategies available when trying to solve others. It is these complementarities and externalities that we are interested in here.

Closely tied to the legal system is the choice of a constitution. For example, the Magna Carta is a seminal document aimed at limiting the strategic choices of royalty and asserting the rights of barons. Countries that base their constitutions on the principles laid out there are able to protect themselves from exploitation by elites and create a Final Game with a legal system more amenable to growth. In the American experience, when our Founding Fathers wrote the original constitution and Bill of Rights they were

[1] Obviously, religion plays an important role in the cultural theories of economic development as Weber (1930) describes.

intentionally binding the hands of future generations about a wide variety of strategies they might use. Although laws can evolve and strategy restrictions reversed, evidence suggests that such reversals are difficult. For example, in determining that state militias could arm themselves against the tyranny of the federal government, the American Bill of Rights restricted our current ability to control guns today, leading to a heavily armed citizenship as opposed to say, Canada or England, whose constitution has no such provisions.

Likewise, in a more contemporary setting, countries in the European Union who adopt the Euro give up sovereignty over their currency. This means that countries, such as Greece, have limited strategies for dealing with economic crises. They cannot control their currency and may be prone to longer and more devastating economic recessions. Finally, again as Acemoglu and Robinson (2012) demonstrate, choices made during a country's colonial experience can have dramatic effects on the Final Game played by that society.

The key point of our paper is that, in the presence of complementarities, choices early on in development (the Formative Games) have a long-lasting impact on the final institutional structure of a society (the Final Game). Those societies that navigate their way successfully through these Formative Games determine efficient institutions while those who don't are left with inefficient Final Games.

In this paper we focus on three factors which we think are important in determining the difficulty different societies face when navigating this institutional maze. These factors are expected to affect the difficulty of the institutional task for different reasons; some behavioral and some theoretical. Each factor determines an experimental treatment whose effects on efficiency we examine.

The first factor is the order in which societies face the Formative Games they do. Though we set up the experiment so that order will not theoretically matter, we expect that it could have effects for two behavioral reasons. In our experiments, we have subjects face three Formative Games: a Trust Game, a Prisoner's Dilemma and a Pure Coordination Game. The way they navigate these games affects the Final Game they eventually play. In one treatment, we have subjects play these games in the order Trust→Prisoner's Dilemma→Pure Coordination, while in another the order is the opposite: Pure Coordination→Prisoner's Dilemma →Trust. Our first motivation for this treatment is the well documented idea that societies that develop a culture of trust tend to have better development experiences. Support for this hypothesis comes from the work of Tabellini (2008, 2010), Putnam et al. (1993), and others where it is demonstrated that morality or higher levels of social capital in the form of generalized trust of outsiders are key inputs for societies in their attempt to solve institutional problems.

For example, in our experiment we might expect that if subjects exhibit trust when playing the Trust Game first, that will help them cooperate in the subsequent Prisoners' Dilemma, an essential outcome for proper Final Game incentives in one of our treatments. Our second motivation is that we believe it is also possible that the distance between the Formative Game and Final Game may matter. For example, when the Trust Game is first, it is three games removed from the Final Game, so we think it is less likely that subjects will be able to look ahead and respond to the existing complementarities. When the Trust Game is placed last, however, it is placed directly in front of the Final Game where the consequences of not trusting may be salient.

The second factor we consider is more theoretically motivated. It is that societies who discount the future less severely will have an easier time in navigating the Formative Games they face. This is true for the obvious reason that many times a successful choice in a Formative Game involves postponing rewards now for the sake of a better Final Game later. Societies that are forced to play Formative Games in times of deprivation (say during times of war or famine) may be forced to make myopic choices simply for short-run purposes which eliminate options that are beneficial in the longer run.[2]

Finally, we also conjecture that some societies face a more complicated set of Formative Games which are harder to navigate, because they involve a more complicated set of actions or complementarities for various reasons such as culture or geography. Here, we define complicated as the number of times the players must give up a myopic benefit to develop the proper strategy set in the Final Game. We demonstrate the latter phenomenon, different complementarities, by comparing our two baseline treatments.

Our contribution is to cast this institutional problem in a game theoretic framework, which we call the Institutional Game, and in the process, stress the complementarities that exist between the Formative and Final Games. In brief, we find strong support for the conjecture that subjects will find it easier to navigate settings where the games they face are easier to solve and where people discount the future less. We find strong evidence that the behavioral complementarities noted above are present, though they work in many ways (e.g. trust leads to cooperation but also cooperation leads to trust) and the Final Games determined are not that different when the order of Formative Games changes.

In this paper, we will proceed as follows. In Section 2 we will describe the underlying dynamic game that forms the basis for our experiments and captures the institutional

[2] One example here might be that of the Danish King Frederick II who, during the Dano-Swedish war of 1660, persuaded his populace to cede him absolute power, a power that lasted until 1848. Perhaps in calmer times the people of Denmark would have not been so eager to abandon a more democratic political institution and give unfettered power to the aristocracy.

problem described above. In Section 3 we will describe our experimental design and in Section 4 we present our results. Section 5 discusses the experiment and its broader implications for societies. Finally, in Section 6 we offer some conclusions.

2 The Institutional Game

2.1 Definitions

We define an Institutional Game as follows. There is a set of N games indexed $t = 1, 2, 3, ..., N$ which are played sequentially by I players. The first N-1 games are called the Formative Games while the N^{th} game is called the Final Game and it differs in the sense that its strategy sets are a function of the strategies chosen by the players in the Formative Games.

More precisely, each of the Formative Games t is represented by its normal form $\Gamma^t[(S_i^t, o_i^t(s^t))_{i=1}^I]$ where S_i^t is the strategy set and $o_i^t(s^t)$ is the outcome function mapping outcomes $s^t \in S^t = S_1^t \times S_2^t \times \cdots \times S_I^t$ to payoffs for player i.

Let $\Phi = S^1 \times S^2 \times \cdots \times S^{N-1}$ denote the set of all histories of play over the $N-1$ Formative Games. The Final Game can be represented in normal form as $\Gamma^N[(S_i^N(\phi), o_i^N(s^N))_{i=1}^I]$ where $S_i^N(\phi)$ is the strategy set which is a function of the history $\phi \in \Phi$ and $o_i^N(s^N)$ is the outcome function for player i. Complementarities are introduced, because the strategy sets in the Final Game depend on the outcomes of the Formative Games.

Future payoffs are discounted by a discount rate δ. As societies develop over a long period of time in some, if not most, cases it may make sense to interpret our game as being played by different generations of players with each player i being a generation who plays one of the N games. Under this interpretation, δ measures the amount by which a generation cares about its descendants.

2.2 An Example

As a clarifying example, and to explain notation for the Institutional Games in the experiment, consider the following four games which represent the $I = 2$ and $N = 4$ Institutional Game played by our subjects in Treatment 1 of our experiment.

Table 1: Institutional Game Example

Formative Games

Formative Game 1		
	C_1	C_2
R_1	17, 17 (-B)	17, 17 (-B)
R_2	4, 39	34, 21

Formative Game 2		
	C_1	C_2
R_1	43, 43 (-A)	9, 60
R_2	60, 9	17, 17

Formative Game 3		
	C_1	C_2
R_1	34, 34	4, 4
R_2	4, 4	26, 26 (-C)

Initial Final Game

	A	B	C	D
A	34, 34	9, 51	129, 9	21, 17
B	51, 9	94, 94	17, 64	30, 30
C	9, 129	64, 17	13, 13	34, 9
D	17, 21	30, 30	9, 34	26, 26

Here we see a set of four games with the first three constituting the Formative Games and a larger game, the Initial Final Game, whose explanation will be made clear shortly. These games are played sequentially starting with Formative Game 1. Note that in the Formative Games the outcomes listed in some of the cells of the matrix contain both payoffs (the numbers incorporate the discount factor) and a "consequence" for the Initial Final Game in the sense that if that cell is the outcome, then a designated strategy of the Final Game will be eliminated. For example, in Formative Game 1, if Player 1 chooses R_1 and Player 2 chooses C_1, then each player gets a payoff of 17 but, in addition, strategy B will be eliminated from the strategy set of each player in the Initial Final Game. This is the sense in which institutions are complementary since the choice of one solution (outcome) in Formative Game 1 both determines an immediate payoff and a consequence for the Final Game to be played by society.

Looking ahead, in Formative Games 2 and 3, we see that there are additional possibilities for other strategies to be eliminated from the Initial Final Game, namely strategies A and C. In total, any combination of the strategies A, B, and C could be eliminated during the play of the Formative Games so there are 8 possible Final Games that may be reached in this Institutional Game. These eight games are summarized in Table 2. We will go into more detail in Section 3 about which of these Final Games are desirable and which are not.

Table 2: Potential Final Games

Initial Final Game				
	A	B	C	D
A	34, 34	9, 51	129, 9	21, 17
B	51, 9	94, 94	17, 64	30, 30
C	9, 129	64, 17	13, 13	34, 9
D	17, 21	30, 30	9, 34	26, 26

Eliminate C			
	A	B	D
A	34, 34	9, 51	21, 17
B	51, 9	94, 94	30, 30
D	17, 21	30, 30	26, 26

Eliminate B			
	A	C	D
A	34, 34	129, 9	21, 17
C	9, 129	13, 13	34, 9
D	17, 21	9, 34	26, 26

Eliminate A			
	B	C	D
B	94, 94	17, 64	30, 30
C	64, 17	13, 13	34, 9
D	30, 30	9, 34	26, 26

Eliminate B,C		
	A	D
A	34, 34	21, 17
D	17, 21	26, 26

Eliminate A, C		
	B	D
B	94, 94	30, 30
D	30, 30	26, 26

Eliminate A, B		
	C	D
C	13, 13	34, 9
D	9, 34	26, 26

Eliminate A, B, C	
	D
D	26, 26

2.3 Final Games and Complementarities

We view the Final Game of our Institutional Game as the game society plays once its institutional fabric has been set. Though in reality this game is played over many, possibly infinitely many, generations we think of the one-shot Final Game as summarizing this interaction in a simple way. The Formative Games, where institutions are developed, are complementary in the sense that they form the rules of the Final Game by constraining the choices available to it.

In our formulation of the Institutional Game, we only allow complementarities between the Formative Games and the Final Game and not among the Formative Games. Many examples, in fact most of the example in the introduction, are of the latter sort. We consider the simplified game, because it helps us isolate the phenomenon we are interested in in a tractable way. In any case, it would be easy to generalize our Institutional Game to allow for complementarities among Formative Games.

3 Experimental Design

3.1 Treatments

We now move to discussing the experiment. A total of 334 subjects were recruited from the general undergraduate population of New York University using the computerized recruitment program of the Center for Experimental Economics at NYU. The experiment lasted a total of 45 minutes and subjects earned, on average, $19.36. The experiment was run using z-tree (Fischbacher (2007)). The payoffs were denominated in experimental dollars and converted to $US at the rate of 10:1.

The subjects in each session were matched into pairs and then played a single $I = 2$ and $N = 4$ Institutional Game like that in the example of the previous section. The subjects only played the Institutional Game presented to them once, because societies just get one chance to develop so the same should be true for our subjects. Since subjects played the game only once, we provided them with a tool to help their understanding of the tradeoffs they faced. The tool allowed subjects to see all the possible Final Games that could be reached after each possible outcome in the Formative Game they were playing if that outcome were to occur. Subjects viewed these Final Games by pressing a button associated with that outcome. For example, using the game from the last section, if they pressed the button in cell (R_1,C_1) in Formative Game 1, they would see the four potential Final Games in which B is eliminated.

To describe our treatments, consider the Institutional Games presented in Tables 3 and 4, each consisting of three Formative Games and the entire set of potential Final Games that could result from some play path of the three Formative Games. Table 3, which combines Tables 1 and 2, provides the games for Treatment 1 (Easy Final Game Baseline) and Table 4 for Treatment 2 (Hard Final Game Baseline). An * denotes the Pareto best equilibrium outcome in each potential Final Game.[3]

The Formative Games proceed as in the example in the last section (they are of course the exact same Formative Games). Now consider the potential Final Games. First, note that the best outcome is (B,B) where payoffs are 94 for each player. We will call this the *efficient outcome.* For some Final Games, (B,B) is an equilibrium while for others, either because it is not available or incentives do not support it, (B,B) is not an

[3]Most of the Final Games have a unique equilibrium but a few have Pareto ranked equilibria. None have multiple Pareto best equilibria.

Table 3: Treatment 1 - Easy Final Game Baseline

Formative Games

Formative Game 1		
	C_1	C_2
R_1	17, 17 (-B)	17, 17 (-B)
R_2	4, 39	34, 21

Formative Game 2		
	C_1	C_2
R_1	43, 43 (-A)	9, 60
R_2	60, 9	17, 17

Formative Game 3		
	C_1	C_2
R_1	34, 34	4, 4
R_2	4, 4	26, 26 (-C)

Potential Final Games

Initial Final Game				
	A	B	C	D
A	34, 34	9, 51	129, 9	21, 17
B	51, 9	94, 94*	17, 64	30, 30
C	9, 129	64, 17	13, 13	34, 9
D	17, 21	30, 30	9, 34	26, 26

Eliminate C			
	A	B	D
A	34, 34	9, 51	21, 17
B	51, 9	94, 94*	30, 30
D	17, 21	30, 30	26, 26

Eliminate B			
	A	C	D
A	34, 34*	129, 9	21, 17
C	9, 129	13, 13	34, 9
D	17, 21	9, 34	26, 26

Eliminate A			
	B	C	D
B	94, 94*	17, 64	30, 30
C	64, 17	13, 13	34, 9
D	30, 30	9, 34	26, 26

Eliminate B,C		
	A	D
A	34, 34*	21, 17
D	17, 21	26, 26

Eliminate A, C		
	B	D
B	94, 94*	30, 30
D	30, 30	26, 26

Eliminate A, B		
	C	D
C	13, 13*	34, 9
D	9, 34	26, 26

Eliminate A, B, C	
	D
D	26, 26*

* indicates Pareto best equilibrium of each potential Final Game

equilibrium. We call games of the first type *efficient Final Games.* For interpretation, we think of societies that develop proper incentives as those that achieve efficient Final Games where (B,B) is the equilibrium outcome.

Table 4: Treatment 2 - Hard Final Game Baseline

Formative Games

Formative Game 1		
	C_1	C_2
R_1	17, 17 (-B)	17, 17 (-B)
R_2	4, 39	34, 21

Formative Game 2		
	C_1	C_2
R_1	43, 43 (-A)	9, 60
R_2	60, 9	17, 17

Formative Game 3		
	C_1	C_2
R_1	34, 34	4, 4
R_2	4, 4	26, 26 (-C)

Potential Final Games

Initial Final Game				
	A	B	C	D
A	34, 34*	111, 13	9, 9	21, 17
B	13, 111	94, 94	9, 129	30, 30
C	9, 9	129, 9	13, 13	34, 9
D	17, 21	30, 30	9, 34	26, 26

Eliminate C			
	A	B	D
A	34, 34*	111, 13	21, 17
B	13, 111	94, 94	30, 30
D	17, 21	30, 30	26, 26

Eliminate B			
	A	C	D
A	34, 34*	9, 9	21, 17
C	9, 9	13, 13	34, 9
D	17, 21	9, 34	26, 26

Eliminate A			
	B	C	D
B	94, 94	9, 129	30, 30
C	129, 9	13, 13*	34, 9
D	30, 30	9, 34	26, 26

Eliminate B,C		
	A	D
A	34, 34*	21, 17
D	17, 21	26, 26

Eliminate A, C		
	B	D
B	94, 94*	30, 30
D	30, 30	26, 26

Eliminate A, B		
	C	D
C	13, 13*	34, 9
D	9, 34	26, 26

Eliminate A, B, C	
	D
D	26, 26*

* indicates Pareto best equilibrium of each potential Final Game

Treatments 1 and 2 only differ by a few payoffs, but these payoffs are key to determining which Final Games are efficient. In the Easy Game, (B,B) is an equilibrium as long as B is not eliminated in Formative Game 1. In the Hard Game, (B,B) is only an equilibrium of the Final Game with the strategies B and D only. This is where our terminology Easy and Hard comes from. The Hard Game is difficult because not only must the players trust in the first game (the Trust Game), but they then must cooperate in the second game (the Prisoner's Dilemma) and coordinate on the Pareto dominated equilibrium in the third game (the Coordination Game) in order to achieve an efficient Final Game.

These two games are dual baselines from which we will make ceteris paribus changes. The treatments are intended to mimic reasons why we think it may be hard for some societies to attain an efficient Final Game and ultimately the efficient outcome (B,B).

The first change is Treatment 3, (Easy Final Game, Random), which introduces discounting. In this treatment, subjects play the same Institutional Game as in the Easy Final Game baseline except that when they reach the Final Game, the game they play is randomly drawn from the set of Final Games determined by the play of other subject pairs in the room. Though this does not map exactly into discounting, we can think of this as a discount factor of 1 for the Formative Games and a discount factor of 0 for the Final Game. In any case, subjects should not care about the Final Game they determine as they will not play it.[4]

Another reason why some societies may navigate their Formative Games better than others, and have better final institutions as a result, is that they play the Formative Games in a different order. To capture this feature we reversed the order of the Formative Games from the baselines for Treatments 4 (Easy Final Game, Reverse Order) and 5 (Hard Final Game, Reverse Order). These sets of games are presented in Table 5 and 6 in Appendix A.

Treatment 6 was a little different. As the results will show, the Hard Game was very hard. In fact, no subject pairs managed to determine an efficient final game in the Hard Game treatment. There are two competing explanations for this. One is that our subjects simply failed to understand that in order to determine an efficient Final Game, they had to eliminate all strategies except B and D. The other is that while they do understand this, the complementarities they face in the Formative Games get in the way of efficiency. To separate these two explanations we ran a control of sorts for the Hard Game. In Treatment 6, the subjects were given the set of 8 possible Final Games for the Hard Game and asked to pick which one they wanted to play. To ensure incentive compatibility, the subjects actually played either their pick or the pick of the subject they were paired with (the computer chose randomly between the two picks).

[4]We also ran two treatments with the actual discount factors .66 and .45 respectively. We are only reporting the two most extreme discount cases in this paper where the results are most stark. The results for the intermediate discounting cases were in between the two treatments reported here.

Additionally, after they completed this primary task, there was a second task in this treatment in which the subjects were randomly rematched and then they engaged in the same Hard Game baseline experiment as in Treatment 2. The idea behind this treatment is that if in the first task of this treatment subjects recognize the efficient Final Game, but then in the second task they fail to make choices in the Formative Games to reach it, then we have even more indication that what is preventing them from being successful in the Hard Game treatment is not their failure to recognize what the efficient Final Game is, but rather navigating the complementarities involved in the Formative Games.

All five of the main treatments employ the same Formative Games: a Trust Game, a Prisoner's Dilemma, and a Coordination Game. There are countless Formative Games we could have chosen for the treatments in our experiment. We chose these three because we believe these games are most similar to the types of Formative Games societies play. For example, the Trust Game can be thought of as a reduced form analysis for the political game of a dictator choosing between keeping power or handing power back to the people. As Acemoglu and Robinson (2012) state, ruling elites often need to trust that if they give up power, their land and wealth will not be confiscated by those they are negotiating with. As we also believe political institutions are paramount to economic development, the existence of our Trust Game is of vital importance. It carries the main consequence for the Final Game as it determines whether the efficient outcome (B,B) will be available to subjects or not.

Our experimental design is summarized in Table 7.

Table 7: Experimental Design

Treatment		Final Game	δ	Order	# of Subjects (Pairs)
1.	Easy Final Game Baseline	Easy	1	Original	60 (30)
2.	Hard Final Game Baseline	Hard	1	Original	58 (29)
3.	Easy Final Game Random Final Game	Easy	≈ 0	Original	60(30)
4.	Easy Final Game Reverse Order	Easy	1	Reverse	56 (28)
5.	Hard Final Game Reverse Order	Hard	1	Reverse	50 (25)
6.	Hard Final Game Control	Hard	-	-	50 (25)
Total					334 (167)

3.2 Theoretical Predictions

In this section we provide a more rigorous analysis of the equilibria for our five main treatments. We employ the standard definition for equilibria in dynamic games, subgame perfection, which in the context of games with finite horizon can be obtained through backwards induction. However, because our games do not all have unique equilibria, there are multiple subgame perfect equilibria. To obtain a unique prediction, we impose one more restriction. As the game is solved backwards, at each node a Pareto best (for payoffs of the entire continuation game) equilibrium is selected if multiple equilibria exist. The Pareto best selection is unique for the games in our treatments, so this provides a unique prediction for the whole Institutional Game. We call this equilibrium the "Preferred Equilibrium" as it highlights how the complementarities work and because we think it provides subjects with their best chance of determining an efficient Final Game.[5] The Preferred Equilibrium of each treatment is outlined in Table 8.

As we can see, in the Easy Game baseline the equilibrium play path predicted is (R_2,C_1) in Formative Game 1, (R_2,C_2) in Formative Game 2 and (R_1,C_1) in Formative Game 3 leading to a Final Game that contains all available strategies for both players (A, B, C, and D) from which (B,B) is chosen. Note that Player 1 trusts in order to keep (B,B) available in the Final Game which offsets the lost payoff of 13 (4 instead of 17) in Formative Game 1. Players cannot cooperate in Formative Game 2, because there is no incentive to deter a deviation as eliminating A does not change the equilibrium of the Final Game. Finally, players coordinate on the Pareto best equilibrium in Game 3, because eliminating C does not change the equilibrium of the Final Game so they might as well get a larger payoff in Game 3.

In the Hard Game baseline (Treatment 2) the equilibrium play path predicted is (R_2,C_1) in Formative Game 1, (R_1,C_1) in Formative Game 2 and (R_2,C_2) in Formative Game 3 leading to a Final Game that contains only two strategies (B and D) from which (B,B) is chosen. This path is more complicated than for the Easy Game, because this is the unique efficient Final Game. The play in the Trust Game is the same as for the Easy Game for the same reason as above. However, play in Games 2 and 3 is different. In the Prisoner's Dilemma of Formative Game 2, players cooperate because a deviation would fail to eliminate A and thus lead to a Final Game where (B,B) is not an equilibrium. This offsets the myopic gain from defection. In the coordination problem of Game 3, players coordinate on the Pareto inferior equilibrium again because taking the larger payoff in Game 3 fails to eliminate C and thus leads to a smaller overall payoff.

[5]This concept is the same as "consistent equilibrium" defined by Bernheim and Ray (1989) for repeated games.

Table 8: Theoretical Predictions

Treatment	Play Path	Final Game Strategies	Final Game Outcome	Payoffs
1	(R_2,C_1) (R_2,C_2) (R_1,C_1)	{A, B, C, D}	(B,B)	(149, 184)
2	(R_2,C_1) (R_1,C_1) (R_2,C_2)	{B, D}	(B,B)	(167, 202)
3	(R_1,C_1) (R_2,C_2) (R_1,C_1)	{A,C,D}	(A,A)	(35, 35)
4	(R_1,C_1) (R_2,C_2) (R_2,C_1)	{A, B, C, D}	(B,B)	(149, 184)
5	(R_2,C_2) (R_1,C_1) (R_2,C_1)	{B, D}	(B,B)	(167, 202)

The Easy Game where the Final Game is random (Treatment 3) has a different equilibrium play path since in Formative Game 1 it is now an equilibrium action for the Row player to eliminate B by choosing R_1. As the row player does not care about the Final Game, they have no reason to trust and keep B available. Hence we have in Formative Game 1 (R_1,C_1) while in the remaining games we get the same equilibrium actions, i.e., (R_2,C_2) in Formative Game 2 and (R_1,C_1) in Formative Game 3 leading to an equilibrium Final Game with strategies A, C, and D and an equilibrium in the Final Game of (A,A).

The Easy Game and Hard Games with Reverse Orders (Treatments 4 and 5) turn out to be strategically equivalent to the baselines in that the outcome of each game is the same. Of course, because the order of Formative Games is reversed, the play path of Formative Games is also reversed.

In addition to the Preferred Equilibrium described here, there are other subgame perfect equilibria in the Institutional Games played in our treatments. Several interesting ones are discussed in Appendix B.

It is important to point out that we do not consider this paper just a test of our Preferred or any other equilibrium concept. Rather we will focus our results on treatment effects (i.e. comparative statics). While is some cases our treatments lead to different theoretical predictions, in most cases the comparative statics of our model predict no changes in outcome. In these cases, we still may expect different outcomes for behavioral reason as we discuss when we look at the impact of changing the order in which games are played.

3.3 Hypotheses

The results section is divided into three parts, each of which discusses one of the three reasons we think complementarities matter for development. We provide hypotheses for these three reasons here.

Our first section considers the complexity of the complementarities. Clearly, this is a comparison of the two baselines, Treatments 1 and 2. As the analysis so far has suggested, we think it is very difficult to navigate the complex complementarities in Treatment 2 where subjects must trust, cooperate, and coordinate. Therefore, our hypothesis is that more efficiency will be achieved in Treatment 1.

In order to formalize this argument, we might consider a hardness or difficulty metric. For instance, we say that Institutional Game A is harder than Institutional Game B if A requires players to more often sacrifice payoffs that are higher in the short-run Formative Game for the goal of achieving an efficient Final Game. More precisely, we define an index for each player that is the number of times the player must play a dominated action or play an action that leads to a Pareto dominated equilibrium in Formative Games to achieve the efficient outcome in the Final Game. In our Easy Game, this index is 1 for the Row Player and 0 for the Column player since, in the Trust Game, the Row player is forced to give up a payoff of 17 for a payoff of 4. After this game, however, no further sacrifice is required for any player. For the Hard Game these indices are 3 for the Row player who must sacrifice payoffs in all three Formative Games for the sake of the preferred equilibrium and 2 for the Column player. Hence, the Hard Game is harder according to this metric, because the indices for the number of sacrifices the players must make are larger. [6]

To the extent that we observe more efficient outcomes in the Easy Game than in the Hard Game, our argument suggests that this will be due to subjects not making the sacrifices in the Hard Game that are necessary. This leads us to Hypothesis 1 (a).

[6] This is just one of many possible metrics. We suggest it, because it highlights our point that it is difficult when players have to sacrifice immediate payoffs for later gains. Another approach to defining an index might be to use a level-k analysis where the level-0 chooses randomly in each Formative Game and higher levels best respond to the behavior of those beneath them. Under this metric, level 1 and higher types choose actions that determine an efficient Final Game in the Easy Game and no levels choose actions that determine an efficient Final Game in the Hard Game. Perhaps this approach would be more appealing if the analysis led to efficient Final Games in both treatments, but at a higher level for the Hard Game.

Hypothesis 1 (a) *There will be more efficient Final Games, efficient outcomes, and higher payoffs in the Easy Game baseline (Treatment 1) than in the Hard Game baseline (Treatment 2).*

Note that we created our Hard Institutional Game by changing the payoffs in the Final Game but leaving all the Formative Games intact. Theoretically, we noted how just these few payoffs make the Hard Game complementarities more difficult as the players must make the right choices in Formative Games 2 and 3 to eliminate the options that lead to these payoffs. Because the Formative Games are identical, if the subjects fail to make the necessary sacrifices as specified by our metric of hardness, they will play the Formative Games the same in both treatments. This leads to a second testable null hypothesis about the difficulty of Hard Institutional Games.

Hypothesis 1 (b) *Choices in the Formative Games will be identical between Treatments 1 and 2.*

The second section considers the discount factor which is a comparison of Treatments 1 and 3. When subjects do not care what Final Game they will determine, because they will not play it anyway, we expect to see very little trust in the first Formative Game.

Hypothesis 2 *There will be more efficient Final Games, efficient outcomes, and higher payoffs in the Easy Game baseline (Treatment 1) than in the Easy Game with a Random Final Game (Treatment 3).*[7]

Finally we investigate the impact of changing the order in which our Formative Games are played. We do this by comparing the behavior in Treatments 1 and 4 and 2 and 5. Despite the fact that these order reversals do not change our equilibrium predictions, we have reason to expect them to influence behavior. Two conflicting forces are at work here. For example, in Treatment 4, where the Trust Game is moved from first to third place, it is placed directly before the Final Game. This should make the consequences of eliminating B very clear as the Final Game immediately follows rather than having to look forward over several Formative Games to see the consequences of ones actions. For this simple reason we might expect more trusting behavior in the Trust Game of Treatment 4 simply because the efficient course of action should be more obvious.

[7] Note that while Hypothesis 1 makes a prediction that contradicts the Preferred Equilibrium theory, since in both the Hard and Easy Games the theory predicts a fully efficient Final Game, Hypothesis 2 is consistent with theory in that the extreme discounting in Treatment 3 ruins the efficient outcome. While this may seem contradictory, Hypothesis 1 is a test of the behavioral hardness index theory while Hypothesis 2 is a test of the standard theory which is why we have stated them this way.

On the other hand, it is very possible that choices in an earlier game may influence choices in later games. For example, if subjects prove themselves not trustworthy in Formative Game 1 in Treatments 1 and 2 (where it is placed first) this may trigger a set of beliefs which lead to negative reciprocity and out-of equilibrium behavior. In addition, in Treatments 4 and 5, where the Trust Game follows the Prisoner's Dilemma, defection in the Prisoner's Dilemma may lead to a lack of trust in the Trust Game which may counter act the fact that the efficient outcome should be more salient there Such behavior is consistent with subjects who have interdependent utility functions (see Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) or who follow intentions-based theories like those of Rabin (1993) and Falk et al. (1997) where non-equilibrium punishments are exacted. These behavioral complementarities may exist and exert an influence independently of the strategic complementarities that are the focus of our Institutional Games.

In the Hard Game, Treatments 2 and 5, the strategic complementarities are richer and the Coordination Game matters as well. In Treatment 2 the subjects must trust, then cooperate and then coordinate while in Treatment 5 they must coordinate, then cooperate, and then trust. While we expect that trust leads to cooperation and vice versa as noted above, it is less clear what behavioral complementarities exist between players who coordinate in a pure Coordination Game and subsequent trust or cooperation since coordinating in a game where interests are aligned says little about how trustworthy a subject is.

Because we do not know the size of these forces, nor do we have a hypothesis for the behavioral complementarities between the Coordination Game and other Formative Games, we will pose the following null hypothesis.

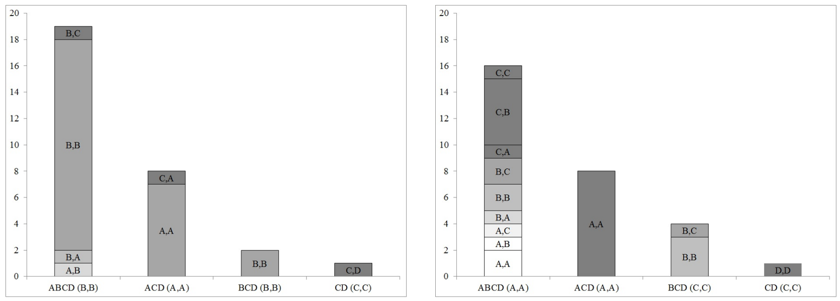
Hypothesis 3 *There will be no difference in the behavior of subjects, the efficiency of the outcomes determined, or subject payoffs when comparing Treatments 1 and 4 and 2 and 5.*

4 Results

4.1 Hypothesis 1

Our first hypothesis concerns the baselines Treatment 1 and Treatment 2. We will begin by describing behavior in these treatments and then compare the results to test the hypothesis.

Much of our interest is in the Final Game, so we will start there. As we know



(a) Treatment 1 (b) Treatment 2

Figure 1: Final Games for Baseline Treatments

from our discussion above, if subjects adhered to our Preferred Equilibrium we would expect that in the Easy Game baseline (Treatment 1) experiment we would observe only the Final Game with all strategies (A, B, C, and D) while in the Hard Game baseline (Treatment 2) experiment, we expect to observe only the Final Game with the strategies B and D.

Of course, in the experiment there is greater variance. Figure 1 displays the distribution of Final Games determined by our subject pairs in Treatments 1 and 2. Also, within each column, the distribution of outcomes in the corresponding Final Game is displayed.[8] For example, in Treatment 1, 8 pairs reached the Final Game with the strategies A, C, and D and seven of these pairs chose the equilibrium outcome (A, A) while the remaining pair chose (C,A).

Looking at Figure 1(a), which focuses on the Easy Game baseline, several results are apparent. Only 63% of pairs reached the Final Game with all strategies still available, the game predicted by our Preferred Equilibrium. However, in the Easy Game, any Final Game that contains the B strategy is efficient since the outcome (B,B) is an equilibrium as long as B is available. Hence, if we add in the 7% of pairs that reached the Final Game with strategies B, C, and D we see that 70% of pairs reached an efficient Final Game.

Additionally, 86% of pairs that achieved an efficient Final Game did indeed end up with the efficient (B,B) outcome. However, taking all pairs into consideration, only 60% of them achieved the (B,B) outcome.

Now let's move on to Figure 1(b) which focuses on the Hard Game baseline. Again there is a variety of Final Games determined but, for the Hard Game baseline, 0% of

[8]The outcome in parentheses in the label under each column is the equilibrium outcome for the corresponding Final Game.

subjects reached the Final Game predicted by our Preferred Equilibrium. As this Final Game is the unique efficient Final Game in this treatment it is therefore the case that no pairs achieved an efficient Final Game. Clearly, the Hard Game was indeed hard.

Considering outcomes in the Final Game, since no pairs achieved the efficient Final Game, we can not analyze behavior conditional on reaching it. However, unlike Treatment 1, the (B,B) outcome is available in Final Games that themselves are not efficient since, while available, (B,B) is not an equilibrium. For such games, only 25% of outcomes were indeed (B,B). Clearly, the availability of the efficient (B,B) outcome in a Final Game is not sufficient for it to be determined. In total, only 17% of all pairs were able to achieve the (B,B) outcome in this treatment.

4.1.1 Hypothesis 1 (a)

In regards to testing Hypothesis 1 (a), there were more efficient Final Games in the Easy as opposed to the Hard baseline treatment (70% vs. 0%) as well as more efficient (B,B) outcomes (60% vs. 17%). Both differences are highly significant (test of proportions, $p < .001$). In summary, with respect to outcomes, the evidence provides strong support for Hypothesis 1 (a).

A second part of Hypothesis 1 (a) is whether the differences between behavior in Treatments 1 and 2 affect payoffs and welfare. Table 9 provides some results regarding welfare for these two treatments (and other treatments we will talk about momentarily) alongside the Preferred Equilibrium welfare.

Table 9: Welfare

Treatment	Eq. Payoffs Final Game	Avg. Payoffs Final Game (s.d.)	Eq. Payoffs Total	Avg. Payoffs Total (s.d.)
1	188	141.40 (59.71)	333	301.97 (70.57)
2	188	106.41 (54.02)	369	256.24 (73.14)
3	188	113.33 (58.47)	333	260.93 (72.94)
4	188	126.50 (63.09)	333	282.93 (79.05)
5	188	86.88 (48.83)	333	249.12 (62.30)

Standard deviations in parentheses

The average Final Game payoff is larger with the Easy Game than the Hard Game baseline, 141.40 > 106.41, as well as over all four games, 301.97 > 256.24. Both differences are significant (t-test, $p = .0218$ and $p = .0176$ respectively). The difference is confirmed by the non-parametric test of distributions (Mann-Whitney test, $p = .0173$ and $p = .0132$ respectively).[9] In summary, the results strongly support Hypothesis 1 (a). The Easy Game determined higher payoffs than the Hard Game.

With respect to welfare, it is also interesting to compare the welfare of subjects that attained efficient Final Games to those that did not to confirm that efficient Final Games were indeed advantageous. Of course, as no pairs achieved efficient Final Games in Treatment 2, this comparison is only possible for Treatment 1. Table 10 compares welfare of those pairs that reached an efficient Final Game, those 70% of pairs that did not eliminate (B,B), versus the other 30% that did.

Table 10: Welfare Difference
Efficient Final Games (Treatment 1)

Final Game	Avg. Payoffs Final Game (s.d.)	Avg. Payoffs Total (s.d.)
Efficient	170.71	338.86
	(43.56)	(44.03)
Not Efficient	73	215.89
	(25.74)	(36.07)

Standard deviations in parentheses

Obviously, payoffs in the Final Game are significantly higher for those pairs that achieved efficient Final Games as they mostly achieved the (B,B) outcome while the others mostly achieved (A,A) (t-test, $p < .001$). More striking however, is that achieving an efficient Final Game was immensely advantageous for overall payoffs as well (t-test, $p < .001$). The difference is not only significant, but very large in magnitude (over 10 dollars).

4.1.2 Hypothesis 1 (b):

Hypothesis 1 (b) investigates behavior at a more disaggregated level by asking why our subjects found it difficult to reach an efficient solution in the Hard Game and what they choose when they did not. First, we want to check whether it is the navigation of the more complex complementarities in the Hard Game that is the stumbling point for

[9]All of the t-tests in this paper have been confirmed by the Mann-Whitney test and we will omit these p-values hereafter.

subjects and not simply that it is hard for them to recognize the efficient Final Game with the implication that if they could recognize it they would navigate the Formative Games effectively. As described above, this is the motivation for the control Treatment 6 and the results of this treatment are positive. When asked which of the eight possible Final Games they would like to play, subjects chose the efficient Final Game with just the strategies B and D 70% of the time. Despite this fact, however, still no subjects determined this efficient Final Game when they played the Hard Treatment. In other words, despite the fact that 70% of the subjects knew they wanted to get to the efficient Final Game, none were able to determine it when they played the three Formative Games before it. They were unable to navigate the complementarities posed by the Formative Games.

Now let's consider how subjects reached their Final Games. Table 11 presents the distribution of outcomes for each of the three Formative Games. Note that the Preferred Equilibrium prediction for each row is a 1, for the predicted choice, and three zeros, for the other choices.[10]

Table 11: Play Path Outcomes

Formative Game	(R_1,C_1)	(R_1,C_2)	(R_2,C_1)	(R_2,C_2)
Treatment 1: Easy Game baseline				
1	.2	.1	.33*	.37
2	.1	.23	.3	.37*
3	.83*	.1	.07	.00
Treatment 2: Hard Game baseline				
1	.17	.14	.24*	.45
2	.17*	.24	.28	.31
3	.59	.31	.10	.00*
* indicates equilibrium outcome				

As we see in Table 11, behavior deviates from the Preferred Equilibrium substantially in both treatments. However, we also see that behavior in the Formative Games seems very similar game-by-game for each of the two treatments (Fisher exact test, $p < .05$ for all three games). The evidence supports Hypothesis 1 (b) that subjects failed to appreciate the complementarities ingrained in their choices and chose to pursue myopic payoffs instead. This, of course, tends to hurt them more in the Hard Game Treatment where there were more complementarities to navigate.

While Table 11 gives a good overview of behavior in Formative Games, it is also interesting to investigate which Formative Games obstructed pairs' determination of

[10]The table denotes with an * the outcomes predicted along the Preferred Equilibrium path of the game.

the equilibrium and/or an efficient Final Game. For this, we look at how many pairs made it through each Formative Game on these respective paths. For the Easy Game, 33% of pairs were on the Preferred Equilibrium path after Formative Game 1, 10% of pairs after Formative Game 2, and 10% of pairs after Formative Game 3. The majority of pairs did not trust with no reciprocation. Of the few that did, the majority of pairs did not mutually defect in the Prisoner's Dilemma. However, of those pairs that stayed on the equilibrium path through the Prisoner's Dilemma, all also stayed on the equilibrium path through the Coordination Game and we eventually had 10% of pairs navigate the game exactly as predicted by the Preferred Equilibrium. In regards to determining an efficient Final Game, as only the first Formative Game determines efficiency, 70% of pairs were on the path to an efficient Final Games after all three Formative Games.

For the Hard Game, 24% of pairs were on the equilibrium path after Formative Game 1, 0% of pairs after Formative Game 2, and 0% of pairs after Formative Game 3. Again, most pairs were off the equilibrium path after the first Formative Game. Even worse than for the Easy Game baseline, for the Hard Game, where cooperation was needed in the Prisoner's Dilemma, no pairs were able to both cooperate. It's impossible to say how difficult coordinating on the Pareto dominated equilibrium in Game 3 was, because no pairs made it that far. We can get a better idea if we look at efficiency. For efficiency, 69% of pairs were on a path to an efficient Final Game after Formative Game 1, 14% of pairs after Formative Game 2, and 0% of pairs after Formative Game 3. For efficiency, the Prisoner's Dilemma was the major roadblock, and those few pairs that made it through couldn't navigate the Coordination Game.

4.2 Hypothesis 2

We now turn to Hypothesis 2 and our second reason for why societies may find it hard to develop efficient institutions. As you recall, in Treatment 3 we theoretically eliminated the impact of complementarities on the Final Game by randomly choosing a Final Game for each pair to play from the set of Final Games determined by the other pairs in the lab.[11] To give an overview of Treatment 3, Figure 2 displays the distribution of Final Games for this treatment in the same way as we showed before for the baseline treatments.

The important difference is that subjects in Treatment 3 only achieve an efficient

[11] In an earlier version of this paper we report the results of other discounting treatments where we gradually changed the discount rate. While behavior did respond to these discount rate changes, it did not do so in a monotonic manner which lead us to use Treatment 3 where it is clear to subjects that their behavior in the Formative Game will have no impact on the Final Game they play.

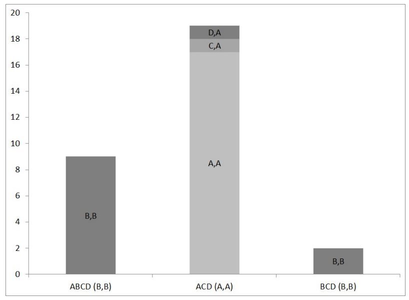


Figure 2: Final Games for Treatment 3

Final Game 37% of the time. This is significantly less than the 70% efficiency reached in Treatment 1 (test of proportions, $p < .01$). When subjects don't care about the Final Game, they achieve efficiency about half as often. In terms of welfare, Table 10 shows that payoffs significantly drop to 113.33 for the Final Games and 260.93 for all four games respectively in Treatment 3 (t-tests, $p = .037$ and $p = .0161$). These results support Hypothesis 2.

In terms of the story we are trying to tell in this paper, the results for Hypothesis 2 suggest that if an event or shock (famine, plague, economic depression) occurs in a society that makes them discount the future more aggressively (or totally) just at the time they are playing a Formative Game with future consequences, then their behavior at that point may have negative consequences for the type of Final Game their descendents are likely to play. This implies that playing the same Formative Games is not necessarily going to determine identical outcomes because societies would need to play them under identical circumstances in order to determine identical results. Adverse shocks of any type that change a society's view of the future may make their behavior more myopic and consequently determine inefficient institutions.

4.3 Hypothesis 3

Before we investigate the effects of order for Final Games, we need to consider the link between cooperation and trust. First, let us concentrate on the behavior of subjects in Formative Game 2 of Treatments 1, 2, and 3 where the Prisoner's Dilemma game follows the play of the Trust Game. More precisely, we consider the behavior of subjects in Formative Game 2 after observing an (R_2,C_2) outcome (where there is trust and reciprocation) in Formative Game 1 as compared to any other outcome (where there is no trust and/or no reciprocation). Table 12, which pools the data of the three applicable treatments, presents these two distributions.[12]

What we see, by looking at the first column in the table, is that there is far more mutual cooperation (R_1,C_1) in the Prisoner's Dilemma after the reciprocation of trust. These results indicate that trust and reciprocation may be a necessary condition for cooperation. This demonstrates a clear behavioral interdependence between Formative Games where an exhibited lack of trust in early games can affect the behavior of subjects later, when they, for example, play a Prisoner's Dilemma game. These results are consistent with a large literature connecting economic performance with cultural values like trust and social capital (see Algan and Cahuc (2014) for a survey). Statistically, comparing Rows 1 and 2, the distribution of choices in Formative Game 2 following trust and reciprocation is statistically different than following other outcomes (Fisher exact test, $p < .01$).

Table 12: Formative Game 2 Conditional Outcomes (Original Order Treatments)

Formative Game 1 Outcome	(R_1,C_1)	(R_1,C_2)	(R_2,C_1)	(R_2,C_2)
(R_2,C_2)	.24	.28	.35	.14
Other	.04	.33	.20	.44

While the results above concern the impact of trust on cooperation in the Prisoner's Dilemma game, in the treatments with the reverse order (Treatments 4 and 5) we have the opposite since the Prisoner's Dilemma game is played before the Trust Game. This allows us to look at the impact of cooperation on trust.

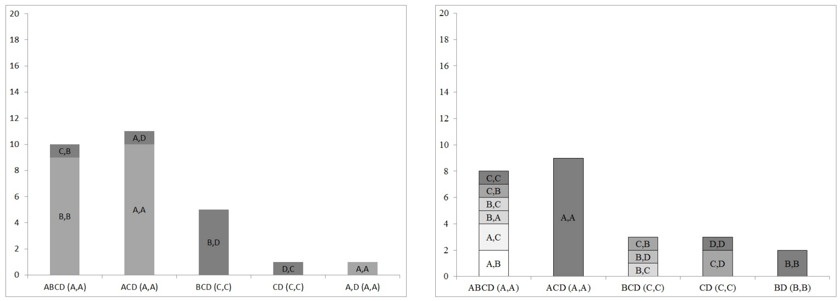
Table 13: Formative Game 3 Conditional Outcomes (Reverse Order Treatments)

Formative Game 2 Outcome	(R_1,C_1)	(R_1,C_2)	(R_2,C_1)	(R_2,C_2)
(R_1,C_1)	.07	.21	.21	.50
Other	.33	.21	.31	.15

Table 13 exhibits a similar behavioral interdependence in this direction. By considering the last column, we see that there is far more reciprocation and trust after cooperation in the Prisoner's Dilemma.

The distributions of outcomes across the first row of Table 13 differs from that of the second row, (Fisher exact test, $p = .054$), indicating that play following an (R_1,C_1) differs from play that follows any other outcome in Formative Game 2.

[12]The trends in the following are generally true for each of the three treatments separately, but the quantity of data is too small to make statistical inferences.



(a) Treatment 4

(b) Treatment 5

Figure 3: Final Games for Reverse Order Treatments

Note that the Preferred Equilibrium in the Easy Game with the reverse order calls for defection in the Prisoner's Dilemma, but then trust in the Trust Game. This is actually quite difficult, as subjects look back at the outcome of the Prisoner's Dilemma when deciding whether to trust or not. The Prisoner's Dilemma spoils the Final Game by exerting a behavioral complementarity on the Trust Game that leads to less trust than when the Trust Game is played first.

There are two possible explanations for the correlation between behavior in the Prisoners' Dilemma and Trust Games. One is that there are behavioral types in the population and for some types trust and cooperation are correlated so if we see trust in one game we are likely to see cooperation in another. The other does not rely on types but considers subjects as conditional cooperators or trustors for whom trust in the Trust Game is conditional on previous behavior in the Prisoners' Dilemma and vice versa. We tend to reject the first explanation, however, since if this was the sole explanation of our data then, when a subject who was more cooperative/trustworthy was paired with one that was less so, the more cooperative/trustworthy subject would continue to cooperate or trust even after a bad outcome. We don't find evidence of this in our data so we emphasize the story of conditional reciprocity. Additionally, this explanation would not account for the difference in behavior in the Easy Game between the regular and reverse order that will be noted shortly. Despite our findings, we believe it is possible that some of the observed behavior could be due to correlated behavioral traits.

We now investigate Hypothesis 3. Put simply, while the order or play did have some interesting consequences, in terms of efficiency or payoffs the results were generally insignificant. We begin with Treatment 4 and focus on Final Games; in particular efficient Final Games though the whole distribution of Final Games in the Easy Game treatment is presented in Figure 3(a). If distance from the Final Game was the primary effect, we would have expected the number of efficient Final Games to increase from the established baseline, where 70% of the Final Games were efficient. The number of efficient Final Games actually decreased to 54%, (test of proportions, $p = .0988$). The fact that about half of Row players did not trust and eliminated B is perhaps surprising given the significant impact that eliminating strategy B has on payoffs and the fact that subjects gave up 60 in the Final Game, for a gain that was at most 13. However, it makes sense in light of the argument above that it is very difficult to trust after defection in the Prisoner's Dilemma. When the Formative Games are played in reverse order, the Prisoner's Dilemma spoils the relationship, and leads to no trust in the Trust Game despite the very large and immediate payoff advantage of doing so.

For Treatment 5, the Hard Game Reverse Order treatment, the one interesting result for changing the order was that 2 pairs successfully navigated the complementarities to get to the efficient Final Game. This shows that the Hard Game is very hard, but at least not impossible. However, as it is only 2 pairs, we are reticent to say that the reverse order of Formative Games made the Hard Game easier than the original order.

From the standpoint of welfare, the evidence in Table 10 points to small losses when the order is reversed for both the Easy Game and the Hard Game. The only (marginal) significance in any of the comparisons is for the comparison of welfare in the Final Game with the reverse order, 86.88, to that with original order, 106.42 (t-test, $p = .086$). It is worth noting that even though the two pairs that successfully navigated the Hard Game were in the treatment with the reverse order, average payoffs across all pairs were actually slightly higher with the original order.

Finally, the order of play appeared to have no impact on the behavior of our subjects in regards to the Coordination Game. In other words, choices in the Coordination Game were invariant to whether the Trust Game was played first or last.

It is important to emphasize why these results are important for the punch line of this paper. Specifically, when we analyze these games using backward induction, at each point subjects are supposed to look forward and envision what will occur there before they choose at any given time. Doing so is supposed to yield, in our games, an efficient result. These behavioral results indicate, however, that some subjects are also looking backward at the behavior of their opponents and either punishing them for their actions or updating their belief about their intentions. Such behavior can lead subjects to take actions that irrationally eliminate strategies in the Final Game, as we saw was the case for the Easy Game with the reverse order, and hence leave their descendents an unsatisfactory Final Game to play.

5 Discussion

One of the main punchlines of this paper is a demonstration of exactly how hard it is for any given society or people to develop an efficient set of economic and social institutions. In fact, our paper offers what might be considered an upper bound on the probability with which groups of economic agents are capable of establishing a set of workable and efficient institutions given the complementarities that exist in Institutional Games.

To understand why we consider our results to be upwardly biased, consider how the Institutional Game has been played historically. As is the premise of our paper, societies face the same types of formative problems we outlined here. They must decide on a political system, a banking system, a religion, etc. However, in the real world, these problems are not solved by individual agents who have perfect knowledge of the set of games they will face in the future and the complementarities between them. Rather, in real societies, generations of social agents are born and interact in the Institutional Game for a limited period of time and then die only to be replaced by others. If during their lifetime an important Formative Game is solved, then that generation's behavior will have an impact on future generations.

Second, solving these Formative Games may be done either by an elite set of agents who are politically powerful or by the spontaneous actions of agents who create (informal) institutions, as Karl Menger said, through their actions but not by their design. In the latter case, each agent has little impact on the outcome of the game but collectively their actions determine a solution that gets passed on to future generations. [13] Acemoglu, and Robinson (2012) would suggest that the key players during this process are elites and that institutions are the outcome of a power struggle between them and those desiring a voice.

Third, unlike our experiment where the Institutional Game is played by agents fully informed about the set of games they will face in the future, in the real world agents have little idea of the future problems they will face, and hence cannot take into account the consequences of their actions on their future choices. In other words, there is no backwards induction taking place. One point of our paper, therefore, is that regardless of whether these societies are oblivious to the interdependencies that exist among Formative and Final Games, these interdependencies are still the factors that determine whether their society will be successful. Whether they know it or not, if they do not make the right choices when playing their Formative Games they will fail to define efficient Final Games. This makes it even harder for societies to succeed.

[13] See Schotter (1981) for a discussion of institutions defined as equilibrium conventions and hence as conventions that are created without a central mechanism designer.

If we define a society as an organized group of persons associated together for religious, benevolent, cultural, scientific, political, patriotic, or other purposes, then there have been thousands or perhaps tens of thousands of such societies that have existed since humans began to walk the earth. We can think of these historical societies as social experiments. Each arrives and faces the Formative Games that need to be faced and either succeeds or fails. Almost all of these experiments ended in failure, however, either with the society ceasing to exist (as say did the Norse community in Greenland between 982 AD and 1500 AD) or by it being absorbed by another. Hence, one way to study societies with functional institutions might be to study their opposite, failed societies, since one might learn what determines success from studying failure.

The study of societal failure has been around for a long time. One only need to look at Arnold Toynbee's famous "A Study of History" to see his analysis of 26 societies that have risen and fallen. In this book, Toynbee (1931-1964), he presents the history of each society in terms of "challenge-and-response". Civilizations arise in response to some set of challenges (perhaps like our Formative Games) of extreme difficulty. These challenges could be physical or social. When a civilization responds to challenges (possibly through the efforts of a "creative minority"), it grows. Civilizations declined when their leaders stopped responding creatively, and the civilizations then sink owing to nationalism, militarism, and the tyranny of a despotic minority.

More recently, Jarod Diamond, in "Collapse: How Societies Choose to Fail or Succeed" (2005), examines a range of past societies in an attempt to identify why they either collapsed or continued to thrive. Among the societies mentioned in the book are the Norse and Inuit of Greenland, the Maya, the Anasazi, the indigenous people of Rapa Nui (Easter Island), Japan, Haiti, the Dominican Republic, and others. While Diamond tends to emphasize ecological factors as responsible for many societal collapses, he does leave scope for political, social and cultural factors of the type examined here. Hence history has offered us a large number of cases where societies fail because they fail to respond to challenges presented by formative problems or games. The more recent book by Acemoglu and Robinson, "Why Nations Fail" (2012), is another continuation of this theme.

We provide a near-perfect environment for success yet even in this environment we observe failure. For example, even in the Easy Game treatment, still only 63% of subjects pairs determined an efficient Final Game while only 70% of the final outcomes were the efficient (B,B) outcome. When we introduced discounting in Treatment 3, this percentage dropped to 37%. In the Hard Game treatment, 0% were capable of achieving either the efficient Final Game.

These results suggest more rationale for our treatments. For example, we ran the

Hard Game treatment to test our notion that an Institutional Game becomes hard when it requires subjects to very often sacrifice short term payoffs for long term gains. It requires coordination across games. While in our experiment such trade-offs could be calculated, in the real world where the set of Formative Games is unknown, it is very likely that people will opt for the higher short-term payoff and therefore play inefficiently. As we saw in our experiment, even in our simple setup, no subject pairs were able to successfully navigate and hence we would expect that such a result would translate strongly to real-world societies. Further, we view discounting as creating a problem for successful institution building in the real world since it forces a myopia or present bias on people leading them to fail to take into account the proper interdependencies that exist (whether they know about them or not). In settings where the set of Formative Games is unknown, this can be an even greater problem.

For all of the above reasons we consider the prognosis for successful institution building as grim. The efficient path that must be navigated by societies is extremely complex and only becomes harder when we rob people of the knowledge of what problems exist in their future and increase the number of agents that exist. The success of real-world societies is basically probabilistic and depends on the ability of decision makers separated by multiple generations to coordinate their actions across the Formative Games they play. This leads us to speculate that one of the advantages of powerful leaders or elites is their ability to better coordinate actions across Formative Games and even choose which problems they want to solve next. Of course, getting them to act in the best interest of society is a moral hazard that has been written about extensively.

Finally our paper suggests a methodological question of whether this type of "history in the lab" is a reasonable way to study institutional development or, more generally, history. We believe the answer is yes. In a typical economic experiment, the elements of economic theories or markets are replicated in the lab and changed in a controlled way to study the comparative-static responses. A similar method can be applied to the study of history; given a theory of historical events one could try to replicate those events or circumstances in the lab and investigate a set of counter-factual exercises in which one would ask a set of "what if" questions to expose alternative histories under an assumption about the motives and preferences of the historical actors.

Furthermore, the experimental methods used here might contribute to helping clarify some of the existing disagreements among scholars interested in institutional analysis where there is a running dispute as to what is the ultimate cause of societal success. While some think it is geography, others focus on culture and yet others on proper institutions. Using the lab to sort out these differences may seem far fetched, but it is not totally so. Looking only at the dispute between culture and institutions, we tend to follow the suggestion of Bisin and Verdier (2016) who suggest that rather than looking for a single ultimate cause, culture or institutions, one might look at the interaction of these two variables and look at situations where cultural and institutional factors interact. [14]

Note that what we are proposing here is different from studying whether people inhabiting different countries or societies play standard laboratory games differently. Such studies have been done often (see for example Roth, Prasnikar, Okuno-Fugiwara and Zamir (1991), Henrich (2000), and Ensminger and Henrich (2014)) We are more interested in whether one can study the creation and evolution of "culture in the lab" and see how this impacts the functioning of economic institutions. Work in this vein has been done by Schotter and Sopher (2003, 2006, 2007) where they look at the creation of equilibrium conventions of behavior in what they call "intergenerational games" used to solve a variety of coordination, trust and bargaining games. In these papers, culture is defined as a set of beliefs and expectations about the behavior of others when confronting recurrent strategic situations. Such beliefs help to support a convention of behavior which governs how people behave when confronted with such situations. Similarly, Camerer and Weber (2003) have looked at culture in firms after a merger while Greif (1994) investigates the institutional set-up of the Genoese and Maghrebi traders and the different conventions they used to regulate trade. We think there is much room to take these cultural and institutional questions to the lab but leave that to future projects.

6 Conclusions

This paper considers the problem of why societies develop differently as most recently articulated by Acemoglu and Robinson (2012). We see the issue as one of institutional development where we define institutions as North (1990) does, the *rules of the game in society*. The question then becomes why do certain societies develop efficient games for their agents to play and others develop dysfunctional ones? We investigate this question theoretically and experimentally in a dynamic game in which subjects play a sequence of games where the solution of one game affects the strategies available in others.

In the experiments we ran, we clearly see the path dependency of behavior and the resulting differences in institutional structures or Final Games. Some societies face more difficult problems than others and typically wind up with less efficient solutions. Others

[14] Experiments that investigate the interaction of culture and institutions have been run by Peysakhovich and Rand (2015) and Schotter (1998). These experiment involve a two-stage design where in stage 1 a culture of trust or cooperation is created among subjects by having them engage in a game known to create (or inhibit) such beliefs and then have them engage in an economic mechanism in stage 2 where their performance is influenced by the conventions or culture created in stage 1.

are more impatient or have lower discount factors and also determine less efficient Final Games. Ultimately, we show how complementarities can lead societies to determine different paths of play and different institutions as they develop.

While we hope that what we have done here has sparked an interest in Institutional Games, we consider this to be only a first step on the road to investigating these games. One obvious extension, both theoretically and experimentally, is to investigate an Institutional Game in which the complementarities were among the Formative Games rather than between the Formative Games and the Final Game. Clearly, as many of our examples indicate, there are many instances of such externalities and we intend to look at them as a next step. However, our focus here was to look at the institutional structure societies inherit from the past and hence we have taken our current path. Exploring only interactions between Formative Games tells a slightly different story.

A second avenue for exploration is modeling the Final Game as an infinitely repeated game, "The Game of Life," which plays out the rules determined in the Formative Games ad infinitum. In our current paper the Final Game is static and may be viewed as a reduced-form version of the true repeated game played by our players. It might be interesting to run experiments where subjects know that what they do in the Formative Games is setting the rules for their repeated interaction in the future. This would place more attention on the Final Game and possibly lead to better (less myopic) behavior in the Formative Games.

Another extension might be to examine a situation where there is uncertainty when playing any Formative Game about the type of game that will arise next and the complementarities it might involve. This would take the emphasis away from backward induction and possibly place more of a premium in making choices that allow for the most flexibility for the future.

Finally, this paper explores the theory of inter-connected games and hence takes a step in the direction of reality since in the real world the games we play are not isolated one from the other. In this paper, we have a sequential structure while in Avoyan and Schotter (2015) there are a set of games played simultaneously under an attentional constraint. Such inter-related games, whether sequential or simultaneous, are clearly important and worth deeper examination.

References

[1] Acemoglu, Daron, Johnson, Simon, and Robinson, James A. Institutions as a Fundamental Cause of Long-Run Growth. in *The Handbook of Economic Growth*

Aghion, P., Durlauf, S. (Eds.), 1A. Elsevier, Amsterdam, 2005.

[2] Acemoglu, Daron, Johnson, Simon, and Robinson, James A. The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review* 2001. 91(5), pp. 1369-1401.

[3] Acemoglu, Daron and Robinson, James A. Why Nations Fail: The Origins of Power, Prosperity, and Poverty. Crown Publishers, New York, 2012.

[4] Alesina, Alberto and Giuliano, Paola. Culture and Institutions *Journal of Economic Literature* 2015. 53(4), pp. 898-944.

[5] Alesina, Alberto and Giuliano, Paola. The Power of the Family, *Journal of Economic Growth* 2010. 15, pp. 93-125.

[6] Algan, Yan and Cahuc, Pierre. Trust, Growth, and Well-Being: New Evidence and Policy Implications. in *The Handbook of Economic Growth* Aghion, P., Durlauf, S. (Eds.), 2A. Elsevier, Amsterdam. 2014.

[7] Avoyan, Ala and Schotter, Andrew. Attention in Games: An Experimental Study. *Mimeo* 2015.

[8] Bernheim, Douglas B. and Ray, Debraj. Collective Dynamic Consistency in Repeated Games. *Games and Economic Behavior* 1989. 1, pp. 295-326.

[9] Bisin, Alberto and Verdier, Thierry. On the Joint Evolution of Culture and Institutions. *Mimeo* 2016.

[10] Bolton, Gary E. and Ockenfels, Axel. ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review* 2000. 90(1), pp. 166-193.

[11] Camerer, Colin and Weber, Roberto. Cultural Conflict and Merger Failure: An Experimental Approach. *Management Science* 2003, 49(4), pp. 400-415.

[12] Diamond, Jarod. Collapse: How Societies Choose to Fail or Succeed. Viking Penguin, New York, 2005.

[13] Ensminger, Jean and Henrich, Joseph. Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective. Russell Sage Foundation Press, New York, 2014.

[14] Falk, Armin, Fehr, Ernst, and Fishbacher, Urs. On the Nature of Fair Behavior. *Economic Inquiry* 2003. 41(1), pp. 20-26.

[15] Fehr, Ernst and Schmidt, Klaus M. A Theory of Fairness, Competition and Cooperation. *Quarterly Journal of Economics* 1999. 114, pp. 817-868.

[16] Fernández, Raquel. Culture and Economics. *in The New Palgrave Dictionary of Economics: Second Edition* Durlauf, S., Blume, L. (Eds.), Palgrave Macmillan, 2008.

[17] Fernández, Raquel. Does Culture Matter? in *The Handbook of Social Economics* Benhabib, J., Bisin, A., Jackson, M. (Eds.), 1. Elsevier, Amsterdam, 2012.

[18] Glaeser, Edward L. and Shleifer, Andrei. Legal Origins. *Quarterly Journal of Economics* 2002. 117, pp. 1193-1230.

[19] Greif, Avner. Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies. *Journal of Political Economy* 1994. 102(5), pp. 912-950.

[20] Henrich, Joseph. Does Culture Matter in Economic Behavior? Ultimatum Game Bargaining Among the Machiguenga of the Peruvian Amazon. *American Economic Review*, 2000, 90(4), pp. 973-979.

[21] La Porta, Rafael, Lopez-de-Silanes, Florencio, Shleifer, Andrei, and Vishny, Robert W. Law and Finance. *Journal of Political Economy* 1998. 106(6), pp. 1113-1155.

[22] Kuran, Timur. The Long Divergence: How Islamic Law Held Back the Middle East. Princeton University Press, Princeton, NJ, 2011.

[23] North, Douglass C. Institutions, Institutional Change, and Economic Performance. Cambridge University Press, New York, 1990.

[24] North, Douglass C. and Thomas, Robert P. The Rise of the Western World: A New Economic History. Cambridge University Press, New York, 1973.

[25] Peysakhovich, Alex and Rand, David G. Habits of Virtue: Creating Cultures of Cooperation and Defection in the Laboratory *Management Science* 2015. 62, pp. 631-647.

[26] Putnam, Robert D., Leonardi, Robert, and Nanetti, Raffaelle Y. Making Democracy Work. Princeton University Press, Princeton, 1993.

[27] Rabin, Matthew. Incorporating Fairness into Game Theory and Economics. *American Economic Review* 1993. 83(5), pp. 1281-1302.

[28] Roth, Alvin E., Prasnikar, Vesna, Okuno-Fujiwara, Masahiro, and Zamir, Shmuel. Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study. *American Economic Review* 1991, 81(5), pp. 1068095.

[29] Schotter, Andrew. The Economic Theory of Social Institutions. Cambridge University Press, New York, 1981.

[30] Schotter, Andrew. Worker Trust, System Vulnerability, and the Performance of Work Groups. in *Economics, Values, and Organization* Ben-Ner, A., Putterman, L. (Eds.), Cambridge University Press. 1998.

[31] Schotter, Andrew and Sopher, Barry. Advice and Behavior in Intergenerational Ultimatum Games: An Experimental Approach. *Games and Economic Behavior* 2007, 58(2), pp. 365-393.

[32] Schotter, Andrew and Sopher, Barry. Social Learning and Coordination Conventions in Intergenerational Games: An Experimental Study *Journal of Political Economy* 2003, 111(3), pp. 498-529.

[33] Schotter, Andrew and Sopher, Barry. Trust and Trustworthiness in Games: An Experimental Study of Intergenerational Advice. *Experimental Economics* 2006, 9(2), pp. 123-145.

[34] Tabellini, Guido. Culture and Institutions: Economic Development in the Regions of Europe *Journal of the European Economic Association* 2010. 8(4), pp. 677-716.

[35] Tabellini, Guido. Institutions and Culture *Journal of the European Economic Associations* 2008. 6(2-3), pp. 255-294.

[36] Todd, Emmanuel. L'Invention de l'Europe, Seuil, Paris 1990.

[37] Todd, Emmanuel. The Explanation of Ideology: Family Structures and Social Systems, Basic Blackwell, New York, 1983.

[38] Toynbee, Arnold. A Study of History. Oxford University Press, London, 1931-1964 (12 Volumes).

[39] Williamson, Oliver. The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting. Free Press, New York, 1985.

[40] Weber, Max. The Protestant Ethic and the Spirit of Capitalism. 1930.

Appendix A

Table 5: Treatment 4 - Easy Final Game, Reverse Order

Formative Games

Formative Game 1		
	C_1	C_2
R_1	34, 34	4, 4
R_2	4, 4	26, 26 (-C)

Formative Game 2		
	C_1	C_2
R_1	43, 43 (-A)	9, 60
R_2	60, 9	17, 17

Formative Game 3		
	C_1	C_2
R_1	17, 17 (-B)	17, 17 (-B)
R_2	4, 39	34, 21

Potential Final Games

Initial Final Game				
	A	B	C	D
A	34, 34	9, 51	129, 9	21, 17
B	51, 9	94, 94*	17, 64	30, 30
C	9, 129	64, 17	13, 13	34, 9
D	17, 21	30, 30	9, 34	26, 26

Eliminate C			
	A	B	D
A	34, 34	9, 51	21, 17
B	51, 9	94, 94*	30, 30
D	17, 21	30, 30	26, 26

Eliminate B			
	A	C	D
A	34, 34*	129, 9	21, 17
C	9, 129	13, 13	34, 9
D	17, 21	9, 34	26, 26

Eliminate A			
	B	C	D
B	94, 94*	17, 64	30, 30
C	64, 17	13, 13	34, 9
D	30, 30	9, 34	26, 26

Eliminate B,C		
	A	D
A	34, 34*	21, 17
D	17, 21	26, 26

Eliminate A, C		
	B	D
B	94, 94*	30, 30
D	30, 30	26, 26

Eliminate A, B		
	C	D
C	13, 13*	34, 9
D	9, 34	26, 26

Eliminate A, B, C	
	D
D	26, 26*

* indicates Pareto best equilibrium of each potential Final Game

Table 6: Treatment 5 - Hard Final Game, Reverse Order

Formative Games

Formative Game 1		
	C_1	C_2
R_1	34, 34	4, 4
R_2	4, 4	26, 26 (-C)

Formative Game 2		
	C_1	C_2
R_1	43, 43 (-A)	9, 60
R_2	60, 9	17, 17

Formative Game 3		
	C_1	C_2
R_1	17, 17 (-B)	17, 17 (-B)
R_2	4, 39	34, 21

Potential Final Games

Initial Final Game				
	A	B	C	D
A	34, 34*	111, 13	9, 9	21, 17
B	13, 111	94, 94	9, 129	30, 30
C	9, 9	129, 9	13, 13	34, 9
D	17, 21	30, 30	9, 34	26, 26

Eliminate C			
	A	B	D
A	34, 34*	111, 13	21, 17
B	13, 111	94, 94	30, 30
D	17, 21	30, 30	26, 26

Eliminate B			
	A	C	D
A	34, 34*	9, 9	21, 17
C	9, 9	13, 13	34, 9
D	17, 21	9, 34	26, 26

Eliminate A			
	B	C	D
B	94, 94	9, 129	30, 30
C	129, 9	13, 13*	34, 9
D	30, 30	9, 34	26, 26

Eliminate B,C		
	A	D
A	34, 34*	21, 17
D	17, 21	26, 26

Eliminate A, C		
	B	D
B	94, 94*	30, 30
D	30, 30	26, 26

Eliminate A, B		
	C	D
C	13, 13*	34, 9
D	9, 34	26, 26

Eliminate A, B, C	
	D
D	26, 26*

* indicates Pareto best equilibrium of each potential Final Game

Appendix B

In addition to the Preferred Equilibrium described in the main text, there are other subgame perfect equilibria in the Institutional Games played in our treatments, a few of which are worth discussing.

While the payoff in our equilibrium is on the Pareto frontier for the Hard Game treatments, it is not for the Easy Game treatments because both players could improve by choosing (R_1,C_1) in Formative Game 2 (cooperating in the Prisoner's Dilemma). Is there any equilibrium that delivers this payoff? The answer is actually yes. Cooperation is supportable in Formative Game 2 if the equilibrium stipulates coordination on the good payoff in Game 3, (R_1,C_1) if (R_1,C_1) occurs in Formative Game 2, but punishes a defection by stipulating the mixed strategy equilibrium in Game 3 if (R_1,C_2) or (R_2,C_1) occurs in Formative Game 2. In this case, deviating in the Prisoner's Dilemma yields a gain of 17 (60 instead of 43), but this is punished by the mixed strategy equilibrium in Game 3 (which is Pareto inferior and thus ruled out by our criterion). The resulting loss in Game 3 of 17.3 (16.7 instead of 34) is sufficient to deter the deviation.

This is the only Pareto dominating possibility in any of treatments, but if one is interested in welfare (the sum of payoffs) then it would be useful to know if supporting (R_2,C_2) in Formative Game 1 (reciprocating trust) as an equilibrium is possible. This is only possible in the Hard Game baseline. There are equilibria in which (94,94) is not achieved for the Hard Game (see below for more), and using one of these as a threat can deter the Column player from not reciprocating trust.

In any case, we don't find these equilibria plausible for many reasons. First, since subjects only play the game once, it would be difficult to expect players to be able to coordinate on a threat for a game yet to be played. Second, these equilibria are not renegotiation proof and require the subjects to believe they will carry out threats and play Pareto inferior equilibria. Finally, for the case of the Pareto dominating equilibrium in the Easy Game, the punishment is only marginally strong enough to support cooperation.[15]

Other than for Treatment 3, are there equilibria where an efficient Final Game is not reached? As just mentioned, the answer is yes for the Hard Game Baseline. The most plausible such equilibria stipulate (R_1,C_1) in Formative Game 3. Even though eliminating C is required to achieve an efficient Final Game, if the other player is choosing R_1 (or C_1), then there is no way to eliminate C and the best response is C_1 (or R_1). We actually think that these equilibria are quite reasonable and present another reason why the Hard Game is hard; the players must avoid bad equilibria. There are no equilibria

[15] In fact there were no pairs who's choices matched the predictions of either of these two equilibria.

in the Easy Game where B is eliminated so the only equilibria of the Easy Game end with an efficient Final Game. This can be easily seen by noting that if such an equilibria existed the row player could deviate to R_2 in Formative Game 1, which results in a loss of at most 13 (17-4) initially but a gain of at least 60 (94-34) in the Final Game.