Defining Institutional Quality - a Game Theoretic Approach

Andreas Ek

2011-11-29

Abstract

This paper discusses definitions of institutions and proposes three different game theoretic definitions of institutional quality. The setting used is a repeated game between an investor and a country. Three definitions of institutional quality are presented: institutional quality as *utility maximization*, as *credible commitment*, and as *length of foresight*. The merits and drawbacks of the different definitions are discussed, as well as their complementarity and compatibility. The relationship between the technical definitions, and an example of an actual empirical measure of institutional quality are also touched upon.

Contents

Abbreviations

IQ - Institutional Quality

NE - Nash Equilibrium

SPE - Subgame Perfect Equilibrium

IMF - The International Monetary Fund

BERI - Business Environmental Risk Intelligence

ICRG - International Country Risk Guide

PRS Group - Political Risk Services Group

List of Tables

List of Figures

1 Introduction

The study of institutional economics has created a large body of literature during the last century, since the term was first used in a 1919 *American Economic Review* article by Walton H. Hamilton.[1] Today, Institutional Economics usually refers to *The New Institutional Economics,* a term coined by Oliver Williamson in 1975, but with its origin traced to Ronald Coase's articles *The Nature of the Firm* and *The Problem of Social Cost.* In contrast to the old institutional economics that rejects neoclassical economic theory, the new institutional economics modify and expand upon neoclassical theory. From being mainly concerned with transaction costs and their role in the economy, institutional economics is now occupied with a wide range of subjects, such as credible commitment, social norms, asymmetric information, modes of governance, bounded rationality, etc.[2]

During the 80's and 90's, what is commonly referred to as *The Washington Consensus*-policies where the standard recipe for economic growth in developing countries. The basic principles behind this were liberalization, privatization, deregulation, and fiscal discipline in order to get well-functioning markets. That liberalization is not always the ideal way towards efficient markets in the absence of the necessary institutional framework, and that other institutions than markets matter for economic development has become more and more clear over the years, partly because of the failure of these policies in some countries and the success of other countries despite the absence of certain liberalization reforms.[3] This is not to say that Washington Consensus-like policies to get well-functioning markets is not important in acquiring economic growth. Rather, that there are many other important ones, such as institutions for conflict management, social insurance, and regulatory institutions.

Different dimensions of what are good institutions are emphasized in different settings. Their role of lowering uncertainty is often stressed; so also in development economics. However, what good institutions are is not always completely clear. Institutions for making markets perform as frictionless as possible need not necessarily be designed in the same way as institutions optimal for society as a whole. There is still room for confusion about what good institutions and institutional quality (IQ) are. Many times, a clear definition of what exactly it

[1] http://en.wikipedia.org/wiki/Institutional_economics, 2011-09-17
[2] http://en.wikipedia.org/wiki/New_institutional_economics, 2011-09-17
[3] Rodrik 2007, p. 16-17

is, is not given. In an attempt to investigate some of the "progress that would be possible by integrating institutional economics with game theory" (in the words of Ken Binmore) IQ will be discussed using game theoretic language and tools.[4] Different game theoretic definitions of IQ are analyzed, mainly from an investor perspective in developing economies.

The paper begins with some basic game theoretic concepts and definitions of institutions in chapter 2. Chapter 3 introduces the game theoretic model that, with some modifications, will be used throughout the paper. Chapter 4 continues with proposing and discussing the actual definitions of institutional quality, their relation to each other and to a practical example of institutional quality of countries actually used. Chapter 5 gives a brief summary of the discussion from the other parts.

2 Theory

2.1 Basic Game Theoretic Concepts and Notation

A very brief overview of some basic game theoretic concepts used is given here. All notation used is also presented. This is repeated in connection to when it is introduced in the analysis. The reader familiar with these concepts may skip this section without losing any context.

A game in its simplest form is defined by a set of agents, N, that each can take some actions, A, leading to an outcome in the set of outcomes, Θ. All agents, $i \in N$ have a preference ordering, $\preceq_i$over the outcomes $\theta \in \Theta$. Often, as is the case in this paper, the preference ordering is represented by utility functions, u_i for each $i \in N$, so that $u_i(a) \geq u_i(a')$ is equivalent to $\theta(a) \succeq_i \theta(a')\ \forall i \in N$, and $\forall a \in A$. The agents have action choice rules, or strategies, $s \in S$, determining how to act. The actions that follows from certain strategies is denoted by a^s.

The two equilibrium concepts that are mainly used in this text are *Nash Equilibrium* and *Subgame Perfect Equilibrium.* The definitions of these are given below. Sometimes, it is explicit which kind of equilibrium that is intended, sometimes it is not. When it is not, it is not central for the discussion exactly what kind of equilibrium that is referred to.

[4] Binmore, 2010, p. 2

Definition 1. A Nash Equilibrium (NE) in the game, G, is characterized by strategies s^*, inducing action profiles a^{s^*} such that

$u_i(a^{s_i^*}, a^{s^*_{-i}}) > u_i(a^{s_i}, a^{s^*_{-i}})\ \forall s_i \in S_i, s_i \neq s_i^*, \forall i \in N.$

Definition 2. A Subgame Perfect Equilibrium (SPE) of the game G is characterized by strategies s^*, inducing action profiles $a^{s^*}|h$ such that $u_i(a^{s_i^*}, a^{s^*_{-i}}|h) > u_i(a^{s_i}, a^{s^*_{-i}}|h)\ \forall s_i \in S_i : s_i \neq s_i^*, \forall i \in N, \forall h \in H.$

2.2 Theories of Institutions

In daily speech, 'institution' usually refers to particular organizations, such as a university, the British Parliament, or the World Bank. There has been a large range of definitions of the term 'institutions' over the years in the social sciences literature, explicit and implicit. Here, we will look closer at three different definitions: the one proposed by Searle (2005), on the most basic philosophical level, with institutions as collective acceptance; the one put forth by Douglas North (e.g. 1991) with institutions as 'the rules of the game' (as opposed to the players in the game); the one suggested by Aoki (2001) with institutions as equilibrium beliefs. These views represent one philosophical, one economic (in a broad sense), and one game theoretic definition. At the same time, they are overlapping yet complementary.

2.2.1 Institutions as 'Collective Acceptance of X'

In the spirit of twentieth century philosophy, Searle takes one step back and instead of trying to answer the question 'What is an Institution', starts out with answering the question 'What can produce institutional facts'. To explain this, he makes use of three notions: *Collective intentionality, the assignment of function,* and *status functions.*

Intentionality in this sense is a much wider notion than what we commonly understand as intending to do something, e.g. intend to attend a lecture. In its philosophical sense, it refers to the directedness of the mind, or the ability of mental states to be directed at something outside the mind in the surrounding world. To intend to do something is one special case of intentionality; other examples are belief, pride, love, perception, or any state of mind concerned with the outside world. Collective intentionality refers to the ability of groups of human beings, two or more, to share these states of mind. Obvious examples are religious sermons where people express their shared faith, soccer players sharing the state of mind of perceiving the activity they partake in as a 'soccer game' that they are trying to win, or actors in a movie sharing the understanding of each other as actors following a manuscript. All these people act individually, but they do so as part of collective behavior.[5]

The *assignment of function* is giving a function to something that does not intrinsically have this function, for example to assign to a stump the function of a stool. The assignment of *status functions* are a special case of assignment of functions where the object (not necessarily a physical object) is assigned a special function attributed by the collective acceptance of a certain status. Without the collective acceptance of this status, the object would no longer have the function it is assigned.[6]

Through the collective intentionality of recognizing the status assigned to e.g. money, or a trustworthy person, *deontic powers* (powers that carry such notions as rights, duties, obligations, commitments, requirements, certifications, etc.) are created. In Searle's view, basically all institutions are matters of deontic power (without the opposite being true; not everything with deontic powers are institutions). Through their deontic power, institutional structures create desire-independent reasons for action, e.g. to have the duty as a citizen to defend your country might make you choose the action of joining the army although you desire much more to stay home with your wife than risk your life in war. This is a key feature of institutions – their ability to make people act in a certain way regardless of if this specific course of action coincides with their current emotional inclination.[7]

> "An institution is any collectively accepted system of rules (procedures, practices) that enable us to create institutional facts. ... The creation of an institutional fact is, thus, the collective assignment of a status function. The typical point of the creation of institutional facts by assigning status functions is to create deontic powers. So

[5]Searle 2005, p. 6.
[6]Ibid p. 7
[7]Ibid p. 10-11

> typically when we assign a status function Y to some object or person X we have created a situation in which we accept that a person S who stands in the appropriate relation to X is such that (S has power (S does A)).“[8]

The definition of institutions proposed by Searle is a very broad one. It includes institutions as different as friendship, traveler's cheques, the American constitution and the IMF. It also works on many different levels; an institution such as the IMF requires numerous other institutions in place to be able to exist. The most basic institution, needed for any other institution, is some means of communication. Without communication (verbal or non-verbal), collective intentionality and collective acceptance would be impossible; for a number of different people to decide to respect e.g. pieces of paper as payment in exchange of goods or services, independent of each other seems like a zero-probability event.

2.2.2 Institutions as 'Rules of the Game'

Douglas North proposes a simpler and more intuitive definition of institutions than Searle. Institutions, in his sense, are 'the rules of the game', or "the humanly devised constraints that shape human interaction".[9] An important distinction, considering that the word 'institution' in common language often refer to entities such as a parliament (a political 'institution'), is that these humanly devised constraints refer to the rules of the game specifically, as opposed to the players of the game, e.g. organizations. An organization such as a parliament or the IMF may be both a player of a game and partaking in creation and upholding of the rules of the game, thereby easily confused with the actual rules of the game. It is important to maintain this separation.

Institution is to be understood in its widest sense, including both formal rules of the game, such as laws and regulations, and informal rules of the game, such as norms and values. Formal and informal institutions are often interconnected and affect each other both ways. The most prominent societal role of institutions in his view is to reduce uncertainty.[10]

---

[8] Ibid p. 21-22

[9] North 1990, p. 3

[10] Ibid p. 4-6

2.2.3 Institutions as 'Equilibrium Beliefs'

One major concern with the definition of institutions as 'the rules of the game' is that the rules are assumed to be exogenously given. They do not give an answer of where these rules/institutions come from. Some have suggested a kind of meta-game in which the rules of other games are determined, basically just pushing the problem one step back as the rules of the meta-game need to be determined somewhere, and so on infinitely. Defining institutions rather as equilibrias, endogenously determined, solves this problem. There have been a number of proponents for this approach over the years. One example of a game theoretic, technical equilibrium approach to institutions is the one proposed by Hurwicz, defining institutions as enforceable, human-made restrictions on players' actions in a game (the game, in turn, defined by the player set, their respective action set, and the set of all outcomes generated by the players' actions). The key term here is that of enforceability, which Hurwicz defines in terms of Nash equilibrium, i.e. institutions can only be institutions if those upholding the institutions benefit from doing so. The definition we will discuss here is the one proposed by Aoki (2001), with institutions as 'equilibrium beliefs'.[11]

In addition to being endogenous, Aoki (2001) lists four other characteristics he believes a good definition of institutions should have:

1. show institutions' function as summarily representing salient features of how the game is played (not to be restricted to or confused with the rules of the game)
2. robustness in a changing environment (up to a certain point)
3. relevance to all players of the game
4. multiple possible (human-made) institutions, given specified natural and technological order.

> "An institution is a self-sustaining system of shared beliefs about how the game is played. Its substance is a compressed representation of the salient, invariant features of an equilibrium path, perceived by almost all the agents in the domain as relevant to their own strategic

[11] Aoki 2001, p. 6-7

choices. As such, it governs the strategic interactions of agents in a self-enforcing manner, and in turn reproduced by their actual choices in a continually changing environment".[12]

This is formalized in a more technical way:

Consider an equilibrium strategy profile in a stationary environment, $\mathbf{s}^* = (s_1, ..., s_i, ..., s_n) \in \mathbf{S} = \times_i S_i$ where S_i denotes the set of all strategies for agent i. Associated with the equilibrium, there is a function, $I_i^*(\cdot)$, for each agent i, that summarizes the information needed for agent i to choose his/her equilibrium strategy, s_i^*.[13]

$I_i^*(\cdot)$ maps $\mathbf{S}$ into the smallest possible dimensionality such that whenever

$I_i^*(\mathbf{s}) = I_i^*(\mathbf{s}^*)$ for $\mathbf{s} \in \mathbf{S}$,

$s_i^*(\phi(\mathbf{s})) = s_i^*(\phi(\mathbf{s}^*))$ (where ϕ is a consequence function, mapping actions/strategies into the outcome space Θ).

Whenever agent i receives the information $I_i^*(\cdot)$, he/she will choose the equilibrium strategy s_i^* regardless of if the actual equilibrium profile $s^*$is played at that moment. $I_i^*(\cdot)$ is all the information needed for agent i to choose the equilibrium strategy, s_i^*; other details of the other agents' strategies are superfluous or irrelevant to i. $I_i^*(s^*)$ is called agents i's summary representation of i's equilibrium path information set. An analogy is made with market prices: There is no need for a potential buyer of goods in a market to fully understand all possible considerations and motivations for possible sellers; all the buyer needs to know are the current market prices to make a buying decision.[14]

Institutions do not come about in a vacuum. The different environments are represented by different environmental parameters. Denote the set of environmental parameters by $\mathcal{E} = \{\epsilon\}$. Assume there exist a continuous equilibrium mapping of ϵ, $s^*(\epsilon)$, to the set of strategy profiles on a connected subset $\mathcal{E}' \subset \mathcal{E}$, and that there is a common characteristic I_i^* is implied by $I_i^*(s^*(\epsilon))$ for any $\epsilon \in \mathcal{E}'$. Suppose also that among all the environment-common but agent specific characteristics I_i^*, there is a characteristic, I^*, common among all the agents $i \in N$, or formally,

$I^*$is implied by $\{s^*(\epsilon), I_i^*\}$ $\forall \epsilon \in \mathcal{E}'$; $\forall i \in N$ [15]

[12]Ibid, p. 197?
[13]Ibid, p. 198
[14]Ibid, p. 198
[15]Ibid, p. 199

Assume further that within a specific subset of environments, there exist multiple, distinct, equilibrium path strategy profiles $\{s^{**}(\epsilon)\}$, $\{s^{***}(\epsilon)\}, \ldots,$ and analogously with the discussion above, summary representations of the equilibrium path information sets and the common characteristics of these, I^{**}, I^{***}. If all of the above assumptions hold, then the common characteristic summary representations fulfill the five desired requirements on institutions. I.e. it is the I^{*}'s that are referred to as institutions.[16]

2.2.4 Relationship Between the Different Definitions

The definitions are quite different in nature. As noted above, the three definitions presented each represent one philosophical, one economic and one game theoretic definition, with the philosophical as the most basic, and the economic and game theoretic in turn. For example to have humanly devised constraints, the most basic institution according to Searle, some means of communication, is needed. In turn, to reach shared beliefs about how a game is played, some rules of this game would be needed.

The most important difference is that institutions with Searle's definition may refer to (but is not restricted to) actual organizations or players of a game, and the other two definitions emphasize the distinction between players of a game and institutions. Searle's definition is most likely the widest, as it contains institutions that the others do not and it is hard to think of institutions according to one of the other definitions that would not be or contain an institution with Searle's definition. 'Humanly devised constraints' on interaction, to the extent that a constraint actually constrains human behavior, should be something that a collective can direct their minds towards and assign a status to that constrains their behavior in a some way, thereby being an institution in Searle's sense. The same goes for Aoki's definition as it presupposes the existence of rules of the game.

Still, being quite different in nature, it is not possible to talk about them as subsets of each other. A set of rules, an institution according to North, is not (necessarily) part of the set of institutions according to Searle, although in the setting containing these rules, there must also exist an institution as collective acceptance of X. The same is true for the relationship 'rules of the game' - 'shared beliefs about how the game is played'; the shared beliefs are not an institution

[16] Ibid, p. 199

but for the institution of shared beliefs to exist, there must be a rules of the game-institution. That Searle's definition is wider is to be understood as that in a setting where one of the other institutions exist, a collective acceptance-institution necessarily exist, without the opposite being true.

How the different definitions of institutions affect definitions of IQ and relationships between these are discussed below in section 4.4.

2.3 Institutional Quality

That institutions matter is no longer seriously debated. The question now is rather which institutions matter the most, and in what way they matter. Many times, IQ seems to be defined loosely as institutions associated with improved economic performance. E.g. Rodrik (2007) talks about "good" institutions as those supporting economic growth in the best possible way. He lists regulatory institutions, property rights, institutions for macroeconomic stability, social insurance, and conflict management, as those that matter most in that respect.[17]

It has also been done explicitly in formalized and very specific ways. Levchenko (2007), for example, take IQ to be the quality of contract enforcement and property rights, captured in a parameter expressing to what extent an investor can get back her/his ex ante investment. A similar definition is given by Klein (2005), where IQ is the extent to which an investor is protected against expropriation. Perfect institutions in this case correspond to zero risk of expropriation.

Chong and Calderón (2000), in an empirical attempt to establish a two-way causal link between IQ and economic growth, uses measures of contract enforceability, nationalization potential, infrastructure quality, and bureaucratic delays from the Business Environmental Risk Intelligence (BERI).

There are a number of different services like BERI available, trying to measure various dimensions of institutional risk and quality in countries around the world, often aimed at big investors. One frequently cited is *The International Country Risk Guide* from the *PRS Group*. It is built around the three dimensions of Political, Economic, and Financial risk ratings, constituted by a number of individual components seen in table 1 and 2 below, and will be discussed in more detail in section 4.5.

[17] Rodrik 2007, p. 154

Political	
Government Stability	Military in Politics
Socioeconomic Conditions	Religious Tensions
Investment Profile	Law and Order
Internal Conflict	Ethnic Tensions
External Conflict	Democratic Accountability
Corruption	Bureaucracy Quality

Table 1: Political Components of the International Country Risk Guide

Economic	Financial
GDP per Capita	Foreign Debt, % of GDP
Real Annual GDP Growth	Foreign Debt Service, % of Exports
Annual Inflation Rate	Current Account, % of Exports
Budget Balance, % of GDP	Net International Liquidity, Months of Import Cover
Current Account, % of GDP	Exchange Rate Stability

Table 2: Economic and Financial Components of the International Country Risk Guide

Exactly which part of institutions and IQ that investors care the most about is likely to vary with the kind of investment, investor and country we are concerned with. The last two definitions at least capture one part of what basically all investors cares very much about.

3 The Model

3.1 Why formalize the definition of Institutional Quality in a Game Theoretic Model?

When discussing complex matters, there is always a risk of getting lost by confusing concepts not clearly defined. As institutions and IQ are quite complex matters, there is a value in clearly defining the precise meaning of these concepts. Using the game theoretic apparatus with its mathematical language is one way of eliminating confusions. The literature in general lack a formal definition of what exactly IQ is, why a discussion concerning what this is in different settings is called for.

Defining IQ by using game theoretic concepts is also a way of studying the logic behind various possible interaction patterns and considerations in situations with conflicting interests of the parties involved. When setting up a game theoretic model of a particular situation, we try to capture the essential elements of some strategic situation. Doing this in an elaborate way may provide insights not obvious at first sight by sorting out possible effects of the agents' intertwined considerations about how to act.

Regardless of whether a definition of IQ can be agreed on at a certain level of consensus, simply having the discussion on what we refer to with institutional quality is worthwhile having for at least two reasons:

1. The process of discussing differences of opinion may generate new insights
2. Even if one clear definition can not be agreed on, there is clear value in having the core of disagreements on precisely defined.

3.2 The Game Setting

Four different definitions of IQ will be proposed and formalized in a game theoretic way. We depart from the different definitions of institutions given above. The definitions of IQ are given and discussed mainly in a particular investor-country game from an investor perspective, trying to answer what exactly IQ is for an investor who considers making an investment in a particular country. To what extent the definitions are generalizable will also be discussed.

Consider the situation of an investor facing the choice of investing in a country or not. The investor cares about the return of the investment, and if he/she gets to keep the return (which of course could be incorporated in the investor's expected return). Let this in reality quite complicated situation be represented by a simple extensive form game with an investor and a country. In period 0, the investor chooses whether to invest in the country or not. This is followed by a number of repetitions of a simultaneous move game where the investor chooses between the actions of committing to the investment and leaving the country, and the country chooses between respecting the investment and stealing, until the investor chooses to leave, or the "end of time". With some adjustments, this is the game that will be discussed.

Formally, let the game (G) be defined by:

- A set of agents or players, $N = \{1, 2\}$, where 1 is to be interpreted as the investor and 2 as the country.

- A set of actions, $A = A_1 \times A_2$; $A_1 = \{Invest, Commit, Leave\}$, $A_2 = \{Respect, Steal\}$. An action by player i in period t is denoted by a_i^t and an action profile (a_1, a_2) in period t by a^t.

- A set of histories, $H = \{h\}$, containing information of what has happened up to a certain point, possibly the end point.

- A set of strategies (or action choice rules), $S = S_1 \times S_2$, assigning actions for the players after any history, regardless if this history will come about or not when the strategies s are followed. A series of actions induced by strategies $s = (s_1, s_2)$ is denoted by a^s, or $a_i^{s_i}$for the actions of player $i \in N$, if nothing else noted over the entire duration of the game from the beginning to the end of time. Actions induced by strategies s after a history h are denoted by $a^s|h$. Actions in period t induced by the strategy s is denoted by $a^{t|s}$.

- A payoff function representing the players preferences over combinations of actions, $u_i(a)$ where i prefers a to a' whenever $u_i(a) > u_i(a')$ $\forall i \in N$, $\forall a \in A$ (and analogously for $<, =$). The utility following after a history, h is denoted by e.g. $u_i(a^s|h)$

- An ordering, or time, of the game, $t \in T$, $T = \{0, 1, 2, ..., n\}$where n may be finite or $\to \infty$

- A player function, ρ, assigning to each $\rho(t)$ one or both of the players to move. $\rho(0) = 1$, $\rho(t) = N \forall t > 0$

The game can be represented by figure 1 below, where $u_i(a_1, a_2) = x_i$, or e.g. $u_i(commit, respect) = r_i$. The table in the bottom right shows the constituent game that is repeated until the end of time or until the investor leaves.

Some notes on possible interpretations of the various actions are needed:

Investing *in* the country, could be interpreted as representing the full cost associated with the activity undertaken by the investor. For example, in the case of setting up a production facility from scratch somewhere, it would be all costs incurred by this new facility; everything from building it and buying the equipment to finding the necessary employees, paying for local licenses, etc. Another possibility is that the investment is moving an existing factory with its equipment and key employees; a third is that it

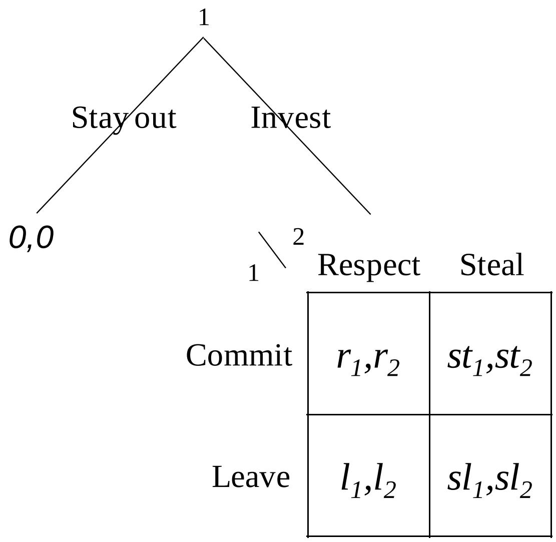

Figure 1: Game Tree Representation of the Investor-Country Game

is allocating some resources already paid for to a yet unspecified country. The interpretation has implications for the relation between the various payoffs. In the first scenario, the country could not steal more during one period than what investing costs the investor, if the investor decides to leave in the same period, $sl_2 \leq I$. Depending on the interpretation of *committing* and *stealing*, st_2 might or might not be able to be greater than I.

Stealing could be everything between taking full control of the investment, its employees and returns, and increasing taxes in a way not anticipated by the investor. Again, depending on the interpretation of *investing* and *committing,* st_1 might or might not be able to be below zero.

Respecting the investor is the country sticking to a stable institutional environment and everything that was expected by the investor, providing a setting in which the investment yields the payoffs hoped for by the investor. This includes not raising taxes in an unexpected way, averting any risk of pillaging, disrespecting patents, stealing resources such as proprietary technology, etc.

Commit could be just staying in the country and maximizing returns without dedicating any additional resources. It could also be thought of as using part of the returns to reinvest in the country, or including a commitment to pay certain wages and fees, making it possible for $st_1 < 0$, even in the situation in which the investment, $I > 0$, symbolizes the most significant part of what could be stolen, and this cost where already accounted for in the *investment* itself.

Leaving the country is basically the investor abandoning the investment, ending the game. Both agents get zero payoffs for all periods after the leaving period. In the *leaving* period, the investor could get a payoff of both greater, equal to, and less than zero: greater if the investor manages to get back part of the initial investment when it leaves; less than zero if the country by *stealing* in this period manages to take something inflicting a loss upon the investor not included in the initial investment; and zero if nothing happens except for the investor departing, or the net effect of the possible effects equals out.

The important thing to notice here is that combinations of (I, st, sl, r, l) that one might instinctively object to, e.g. $st_2 > I$, or $-st_1 > I$ may be possible with a certain combination of interpretations. This highlights the importance of clearly defining

There are a number of trivial situations (mainly with somewhat strange interpretations of what the actions represent) depending on the orderings of the outcomes (e.g. if steal and leave, respectively, are strictly dominated actions by respect and commit). We will look at the non-trivial situation when $r_1 > l_1 > sl_1 \geq st_1$ and $st_2 \geq sl_2 \geq r_2 > l_2$, $st_2 > r_2$. Looking at this as a one-shot game, steal strictly dominates respect for the country, why the country playing steal and the investor playing leave is the unique NE (unless $st_1 = sl_1$, when Commit, Steal is also an equilibrium). If the game instead is repeated, the situation may differ.

3.2.1 Repeating the Game Finitely

We continue by introducing time, where at each point in time, the constituent game is repeated. To start with, time (n) is finite. Then, there is a last period of the game where (if this period is reached) player 2 will play *steal* and player 1 will play *leave*. Assume there is a $k < n$, $k \in T$, for which $(steal, leave)$

will be played all periods (k, $k+1$, ..., n). If so, since the investor knows that the country will play steal in all periods following $k-1$, giving the investor a payoff of less than zero, the investor will prefer to leave already at time $k-1$. Knowing this, the country prefers to steal; hence (*steal*, *leave*), will be played at time $k-1$ as well. Looking at $k = n-1$, as the investor knows that the country will play steal next period, the investor will leave and the country steal at period $k = n-1$. As $k = n-1$ is less than n, the condition above holds. That is, there is a $k < n$ for which (*steal*, *leave*) will be played all periods (k, $k+1$, ..., n), and (*steal*, *leave*) will be played at $k-1$ as well. As it is true also for $k = (n-1)-1 = n-2$, it must be true also for $k = (n-2)–1 = n–3$, and so on for every period $t \in T, t > 0$. By induction, we see that the unique NE (and SPE) is that the investor never invests in the country, and (off the equilibrium path) (steal, leave) is played every following period.

Note that this is true regardless of any (possible) beliefs of the agents concerning strategies of the other agent. For example, one might think that it could help if the investor had a formulated strategy of investing and committing as long as the country respects, and leaving once the country steals, but not before; the country, knowing about this, and preferring $\sum_{t=1}^{k} r_{2,t}$ to 0, for some $k > 1$, would not steal during k periods. The problem is that it is impossible for two rational agents, and common knowledge[18] about this rationality, to sustain any such beliefs of that a strategy like that would be followed. Since the investor knows that there is a last period where the rational country steals, and the country knows that the investor knows this, the country will take the opportunity to steal already in the second to last period. And since the investor knows that the country will do this, the investor will leave in the third to last period, the country will steal in the fourth to last period, and so on. The country knows that a rational investor cannot follow a strategy such as the one described, knowing that the investor knows that the country will (by backwards induction) take the first opportunity to steal. The country cannot credibly formulate a strategy such as the one described, as both parts know that a rational player with knowledge about the rationality of its counterpart would not do such a thing, and common knowledge of rationality was assumed. Agents are trapped by the rules of the game and their common rationality.

As investments after all takes place, there must be numerous situations where

[18]Rationality and common knowledge are to be interpreted in their standard game theoretic sense. See e.g. Osbourne and Rubinstein, chapter 1 and 5 for discussions of these

this description does not capture the agents' perception of the relevant factors of the situation. At the same time, every investment made in one country is also not made in every other country. I.e., we cannot conclude that it does not capture something relevant where the investment is not made. Still, as long as investments are made, there are countries that differ from this description in the rules of the game or institutional setting and quality.

3.2.2 Repeating the Game Infinitely

If we look at almost the same game, but with time, $t \in T = \{0, 1, 2, \ldots, n\}$, $n \to \infty$, i.e. once the investor invests, the constituent game is repeated until the investor leaves, possibly infinitely many times. Instead of one unique equilibrium, the game now has an infinity of equilibrias, of which investor not investing and (steal, leave) played every time is just one. In line with the folk theorems, any outcome where the investor never leaves, leading to a payoff equal to or higher than the minmax payoff[19] of both players can be supported by equilibrium strategies. To see this, consider strategies $s = (s_1, s_2) \in S$ for the two players, leading to payoff profiles $u_1(a^s)$, $u_2(a^s) > 0$. If a player i deviates from her/his strategy in a way that lowers the payoff of player $j \neq i$, j will instantly punish i by playing minimizing i's payoff. In both cases, it will lead to a termination of the game (if the investor deviated, he left; if the country deviated, the investor will punish by leaving). The game is not equivalent to the repeated games in the folk theorem formulations as it is not necessarily repeated infinitely. Looking at the situation once the investor has invested, it could be seen as an infinitely repeated two-player game where one of the agents, if minmaxing/punishing the other, has to do so forever (as leaving is the same as holding down payoffs of both players to zero forever).

Although the number of possible equilibrias is infinite, there are some restrictions on the structure that the equilibrias will take:

- In any equilibrium in which the Investor invests, the sum of payoffs must be greater than or equal to the initial investment I, $\sum_{t=1}^{n} u_1(a^t) \geq I$
- The game must be repeated forever. This is because any strategy with a leaving period $k < \infty$ would function as a last period in which the

[19] The minmax payoff of player i is the lowest payoff that player(s) $-i$ can force player i down to

country would want to steal. As the investor is already leaving in period k, it has no way of punishing deviant behavior in period $k-1$, and the country will steal also in this period. Since $st_1 < 0$, the investor will now leave in period $k-1$, and the situation is back in the finite horizon one (if we allowed $sl_1 = l_1$, the situation could be different depending on the relationship between the different payoffs for the country).

- Assuming that randomization is not allowed and that strategies cannot be infinitely complex[20], any strategy where the investor does not *leave* must have a periodic structure. That is, after some finite number n, the strategy will repeat itself. If the players' strategies lead to repeating their actions after time m and n respectively, the longest possible time before the combination of actions repeats itself will be $n \cdot m$, as after this number of time periods have passed, both players will be "back to the start" of their strategies. The set of equilibrias where the stage game is repeated forever will hence have the structure of a periodic sequence of (*commit*, *respect*) and (*commit*, *steal*) of finite length repeating itself an infinite number of times. The combination of (*commit*, *respect*) and (*commit*, *steal*) over a sequence $n \cdot m$ periods must be such that the sum of payoffs for the investor over each sequence is positive, and the sum of sequence-sums greater than the initial investment.

The two cases of finite and infinite time represent a somewhat binary perspective, where agents either know with certainty that a certain period is the last period, or are completely sure that the game will never end. A less black-and-white view of the world could be to allow for the game to end each period with a probability ϵ (possibly with ϵ as a function of some other variable, e.g. t). In this scenario, both agents know with certainty that the game is finite, as $(1-\epsilon)^n \to 0$ as $n \to \infty$; the game continuing forever is a zero-probability event. Still, the agents never know an individual point in time to be the last and will take into account that the game will continue at least n more periods with probability $(1-\epsilon)^{(n-1)}$(or exactly n more periods with probability $(1-\epsilon)^{(n-1)} \cdot \epsilon$). Even though they know the game to be finite, there is not a point in time from which to do the backwards induction that trapped them in the non-favorable scenario of (steal, leave) always being played. Strategies were the investor invests and commits as long as the country respects, and the

[20] Strategies of finite complexity are strategies that can be implemented by a machine with a finite number of states. See e.g. Osbourne, Rubinstein (1994), chapter 9 for a discussion.

country always respects, are possible in equilibrium, as long as, for the investor, $\sum_{t=0}^{n}(1-\epsilon)^t \cdot r_1 > I$ and for the country $\sum_{t=0}^{n}(1-\epsilon)^t \cdot r_2 > st_2 + sl_2$ for some $n > 0$ (possibly $n \to \infty$). The starting period in the sum, 0, could be any period in the actual game greater than 0, as once the investor has invested, the situation is the same in every period. In this case, it is not in the interest of either agent to deviate, as doing so will yield the lower payoff of 0 or $st_2 + sl_2$ instead of $\sum_{t=0}^{n}(1-\epsilon)^t \cdot r_i$.

In general, the equilibrias in this setting will be the same as in the infinite variation of the game, as long as the expected value of following equilibrium strategies of the infinitely repeated game fulfill the requirements that the certain payoffs do in the infinitely repeated game (assuming risk-neutrality; else, risk aversion would need to be adjusted for, lowering the maximum possible stealing frequency in equilibrium additionally). This is in line with the folk theorem variations in Fudenberg and Maskin (1986), although their proof does not apply here as the investor cannot punish for a limited period of time, there is not (necessarily) any outcome of the constituent game that Pareto-dominates the NE of $(steal, leave)$, and the game is not necessarily repeated more than an arbitrarily small number of times.

As we have seen, the set of possible equilibrium outcomes differs between the two cases of finite and infinite time, with the case of which the final period will be as uncertain somewhere in between, leaning towards the infinite case. Which representation that is closer to the perceived economic reality of the players is not completely clear. There are good reasons to choose either representation of reality, depending on the interpretation of the game. Consider for example the situation of an oil company contemplating whether to invest in drilling for oil on the territory of some country. Even if the company would be extremely lucky, finding an oil well with deposits greater than of any oil well known today, the deposits would still be finite, and the time horizon of the investment decision clearly finite. This is true not only for the physical reality of the situation, as will be the case with any decision on our planet, e.g. by the finite time our sun will continue to heat our planet enough for our life form to exist here. It is also true in the sense that a last period most likely will enter explicitly in the actual assessment that the company will make of the opportunity in a way that the, albeit real, realization that 'I will some day die' do not enter explicitly in the vast majority of decision considerations. Neither will it estimate possible benefits as if the last period when the oil well is estimated to be dry, with some probability would not be the final period; a finite horizon representation seems to be the best suiting way of modeling the situation.

On the other hand, the reason that we get one unique (and not very favorable) equilibrium of the game with a finite horizon comes with the existence of a last, final, 'nothing-beyond-this-point-in-time', period, and the backward induction from this period. If the perception of a situation is such that, even if there in reality is a last point in time, there is no period perceived as a last one, the perception of the situation changes in a way that makes the infinite-horizon representation of the game a more accurate one. (Mathematically, any sequence of numbered elements such that after any element k, there always exist another element $k+1$ is indeed an infinite sequence.) In the case of an investor deciding whether or not to invest in a country, and the country deciding whether to respect this possible investment or not, the country need not necessarily to see the investment as an isolated occurrence. Rather, the 'Investor' could be seen as a pool of possible investors, and respecting or stealing by the Country, is seen as behavior in general, true not only for the isolated investment. Even if there indeed was an explicit horizon for a particular investment, there does not have to be a clear final period for the pool of possible investments; most likely, there is not.

Countries (and in many cases investors) are not individuals but still made up of individuals, why the reasoning of individuals most likely have an impact on how the entities reason. People in general tend to be reluctant to model situations as if there were a final period and draw the backward-induction-conclusions from the existence of this final period. In lab versions of repeated prisoner's dilemma, people often do not use backward induction to play the unique SPE until they are close to the end.

Another slightly less clear example is 'the game of life' where, as noted above, the realization that the life will end does not normally enter into peoples' considerations about their life in the form of 'time period x will be my last'. Even when getting to an age at which any day could, presumably, be their last, most people will live their days as if the probability of today to be the last were negligible and there were at least one more day after the current one. By this line of reasoning, if an agent beliefs that there is a coming period after the current that is affected by the agents' actions in the current period, it should be more accurate to model the situation with an infinite-horizon game than with a finite.

4 Definitions of Institutional Quality

As we try to answer the question of whether or not an investor will decide to invest in a country or not, all the definitions are given from an investor perspective; how the institutional setting and quality is perceived by the investor. Good IQ in this sense might be very favorable for the investor, but perceived as far from optimal from the perspective of the country, the other agent in the game. We begin by introducing the different suggested definitions of institutional quality, discuss their compatibility with each other, very briefly discuss some suggested generalized definitions of institutional quality, and end with an example of an actual used measure of IQ and its relationship and compatibility with the proposed definitions.

4.1 Institutional Quality as Utility Maximizing Institutions

We start with a definition of IQ mainly concerned with efficiency and utility-maximization that could be seen as more or less implicitly given by a number of authors previously, e.g. by North in discussing *institutions **for** economic performance*.

Definition 3. With institutions as 'the rules of the game', IQ from an investor perspective is to what extent the game provides a set of equilibrium opportunities with the highest possible arithmetic mean of total utility derived by the investor in the different equilibrias. Formally, let institutional settings of an investment decision or investor-country interaction be modeled in the games G, G', G", ... The games share the same physical/technological constraints and possibilities, but differ in their respective 'humanly devised constraints'. Denote by E^G, $E^{G'}$, $E^{G''}$, ... the set of equilibrias in the respective games. In each equilibrium outcome $e \in E^G$, agent $i \in N$ receives utility $u_{i,e}$. The statement "G has higher IQ than G' " is equivalent to $mean_{e \in E^G}(u_{i,e}) > mean_{e \in E^{G'}}(u_{i,e})$.

Rules here should be interpreted as everything about how the game is defined, once it is known which situation it is that should be modeled. Once we know that the situation concerns whether or not an investor should make an investment in a particular country, from here on, everything in how the game is set up are part of the rules. The two different situations above, with a fixed and infinite time horizon (denote by G^{fin}andG^{∞} respectively), are two distinct cases of different 'rules of the game' providing very different equilibrium opportunities. As the arithmetic mean utility of the equilibrias in $G^{\infty}$is higher, the IQ is higher in G^{∞}than in G^{fin}. Hence, IQ is defined in a relative, rather than an absolute, sense. A statement such as "the IQ in G is high" is meaningless; if instead one says that "the IQ in G' is higher than in G", the statement carries some economically relevant information.

This is a very inclusive definition, capturing basically any economically relevant alteration of a situation as a change in IQ, except for changes in physical/technological constraints or possibilities. If the country above manages acknowledge that it is a better approximation of their economic reality that in any period of time, there is a tomorrow, and this tomorrow is affected by their actions today, this is an improvement in IQ; when the country lowers labor taxes, or over time induces a stronger work ethic, thereby effectively lowers the relative cost of labor, IQ is improved; when the country voluntarily signs up for some kind of punishment for stealing and thereby removes a part of the equilibrium set with lower total utility, the IQ is improved; etc. These are all things that most would not mind very much to consider improvements of the humanly devised constraints, i.e. institutions, in a country.

There are other things included by this definition that some might object to calling IQ. If for example labor is plainly cheaper in country A than in country B, ceteris paribus, the return of a respected investment and the IQ in A is higher. Calling cheap labor IQ, or lowering wages an institutional improvement does not quite coincide with our general or intuitive understanding of the terms. However, this need not be a very troublesome objection to the definition; one might simply interpret this as the labor market institution being more investment friendly in the case of cheap labor. Generally, it should be the case that countries with cheap labor are in greater need of investments.

Another objection to the definition could be that even though $mean_{e\in E^G}(u_{i,e}) > mean_{e\in E^{G'}}(u_{i,e})$, agents in G' could very well end up choosing strategies rendering the investor higher utility than in G, regardless of the mean of possible equilibrium payoffs. That is, the definition only takes into account the set of possible outcomes, not the actual outcome, which is what the investor cares most about. This brings us to the corresponding institutional-quality-as-utility definition with institutions as a self-sustaining system of shared beliefs about how the game is played:

Definition 4. With institutions as 'a self-sustaining system of shared beliefs about how the game is played', IQ is to what extent the shared beliefs converge towards that the game is played in way maximizing the utility of the investor. Or, using notation as Aoki, $I^*$is an institution of higher quality than I^{**}if and only if the associated strategies $s^*(\epsilon)$ and $s^{**}(\epsilon)$ are such that $u_1(s^*) > u_1(s^{**})$.

Where definition 3 only takes into account actual binding constraints that comes from the rules of the game, definition 4 also takes into account the beliefs of the agents about how to act. It allows for a greater impact on the IQ by the actors themselves, which is natural given that it is a definition of institutions endogenous to the model in contrast to institutions as exogenous rules of the game. In this way, definition 4 is complementary to definition 3: where definition 3 captures what is possible given some kind of 'outer' setting (exogenous rules), definition 4 takes into account what will actually happen as the agents' beliefs about the effects of the rules on the actions by the other agents. One might think that given knowledge about the IQ according to definition 4, an agent would no longer care about that of definition 3, since it already have knowledge of which payoff it will get. However, as the environment, captured by the parameter, ϵ, changes, the institution and IQ may also change. Because of this, IQ as in definition 3 could still be important to an investor as a measure of within which boundaries the outcome may change, even if the IQ from definition 4 is known.

A more problematic objection to these kinds of game theoretic definitions of IQ is that they are simply too inclusive to provide any real insight. That an investor cares about their possible and actual payoff says little or nothing about what can be done to improve the institutional framework, or logic behind interaction patterns not already quite obvious.

4.2 Institutional Quality as Credible Commitment

In much of the literature on institutions, their role of lowering uncertainty is emphasized.[21] This fits well with the institutional definition given by Aoki, and the investor-country-interaction with infinite time horizon discussed above. In that setting, a country would want to (credibly) signal to the investor, their commitment to respect the investor, thereby lowering the uncertainty about how the game will be played out. A problem with the definitions above is that they

[21] E.g. North 1990

are too inclusive get any insight in how institutions affect actions and beliefs of the agents, which kinds of institutions that matter, to what extent they matter, etc.

Definition 5. IQ is to what extent the country is able and willing to send a credible signal to the investor about which equilibrium strategy, s_2^*, the country will use.

To define IQ like this, we have to fall back on Searle's definition of institutions, as it is able to refer to the agents as institutions (in the sense that they can create 'institutional facts') in a way that the other two definitions of institutions are not able to. It does not quite make sense to define an institution to be a 'self-sustaining system of shared beliefs' and refer to the quality of institutions to be a single agent's ability to influence these shared beliefs.

We adjust the games G^{fin} and G^{∞} as follows:

- A probability function π_i expressing agent i's beliefs about which strategy the other agent(s) will play is introduced. π_i assigns probability $\pi_i(\mathbf{s_{-i}}) = p$ to each subset $\mathbf{s_{-i}} \subseteq S_i$. In case a subset $\mathbf{s_{-i}}$ includes more than one possible strategy s_{-i}, i has no further knowledge of which of these strategies that are more or less likely, and assigns the same probability to all strategies $s_{-i} \in \mathbf{s_{-i}}$, $\pi_i(s_{-i}|\mathbf{s_{-i}}) = \frac{1}{|\mathbf{s_i}|}$ $\forall s_i \in S_i$. The agents know which strategies other agents assign positive probability, without necessarily knowing how large probability that is assigned to each individual strategy s_{-i}. $\pi(\mathbf{s}) = \pi_1(\mathbf{s}_2) \cdot \pi_2(\mathbf{s}_1)$.
- The game ordering is extended backwards with a period -1 in which player 2, the country, may send a (possibly costly) signal, σ, adjusting agent $i's$ beliefs from π_i to π_i^{σ}, a probability function with lower variance.
- Any deviation from assigning the same probability to all possible equilibrium strategies is attributed to signaling from the country. Or put differently, any reason the investor has to assign greater probability to strategy s_2 than to any other strategy $s_2^{'}$ is interpreted as caused by some kind of signal from the country.

The game can now be represented by figure 2 below:

Without putting any initial restrictions on σ, the game has the same ex ante set of possible outcomes. As we assume rationality of agents and common knowledge

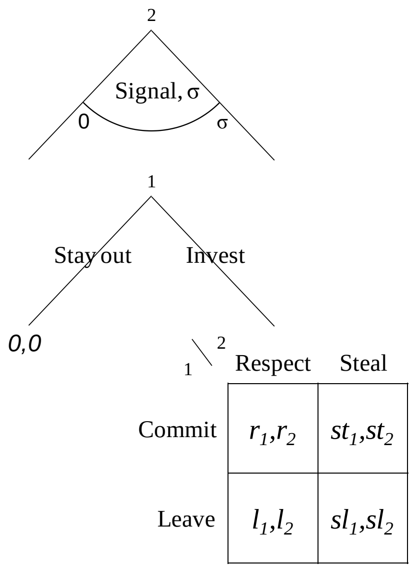

Figure 2: Game Tree Representation of the Signalling Modified Investor-Country Game

of rationality, it is only equilibrium outcomes that are possible. Restricting ourselves again to non-stochastic strategies of finite complexity, there is a finite set of possible equilibrium total payoffs in a game G, call it u_{i,E^G}. The difference in variance of possible equilibrium payoffs before and after σ is given by

$$Var(u_{i,E^G}|\pi_i) - Var(u_{i,E^G}|\pi_i^\sigma) = \sum_{s_{-i}\in S_{-i}} \pi_i(s_{-i})\cdot(u_{i,e^{s_{-i}}} - \mu)^2 - \sum_{s_{-i}\in S_{-i}} \pi_i^\sigma(s_{-i})\cdot(u_{i,e^{s_{-i}}} - \mu^\sigma)^2$$

where $u_{i,e^{s_{-i}}}$ is the utility agent i receives in the equilibrium where $-i$ plays s_{-i} and agent i its best response. μ (μ^σ) is the expected value of u_{i,E^G} given π_i (π_i^σ).

Definition 6. IQ of a country from the perspective of agent i (the investor) is to what extent it is willing and able to eliminate variance of possible equilibrias. The IQ of country 1 is greater than that of country 2 if and only if

$$\frac{Var(u_{i,E^{G_1}}|\pi_i) - Var(u_{i,E^{G_1}}|\pi_i^{\sigma_1})}{Var(u_{i,E^{G_1}}|\pi_i)} > \frac{Var(u_{i,E^{G_2}}|\pi_i) - Var(u_{i,E^{G_2}}|\pi_i^{\sigma_2})}{Var(u_{i,E^{G_2}}|\pi_i)}$$

where the investor-country interaction is described in game G_1 and G_2 respectively.

The IQ of a country is improved if and only if its signal σ is changed to σ' such that

$$Var(u_{i,E^G}|\pi_i) - Var(u_{i,E^G}|\pi_i^{\sigma}) < Var(u_{i,E^G}|\pi_i) - Var(u_{i,E^G}|\pi_i^{\sigma'})$$

To avoid the obvious objection that a country with considerably lower actual variance in of equilibrium payoffs could end up being judged having inferior IQ because the initial variance was lower, we may interpret any difference in initial variance as caused by natural, geographical and technological differences, and everything humanly devised variance decreasing as part of the signal and thereby the IQ.

As the definition here rests upon a single agent's ability and willingness to signal commitment to a particular equilibrium, it is hardly compatible with North's definition of institutions, since it ignores other rules except the signal σ and thereby a large part of the institutional framework. There are compatibility problems with Aoki's definition of institutions as well. Even if the definition was re-phrased with 'a country's contribution to IQ' rather than actual IQ, there is still the problem that in Aoki's definition, agent's information about how to the game is played is part of the institution itself. In order for there to be an institution that can vary in quality, there is no longer any uncertainty about how the game is played, except that caused by shifts in the environment. And if the environment causes an institutional change, it is a new institution, rather than an old one varying in quality. The definition would need to be re-phrased as to signal commitment to a certain course of action also in the event of environmental changes.

Definition 5, however, works well with Searle's definition of 'institutions as collective acceptance of X'. The country needs to credibly assign to itself the status (assignment of status function) of being trustworthy, and in a way that makes investors also accept its status of being trustworthy (collective intentionality of accepting of the function). IQ here corresponds to to what extent the country is able to convince outsiders, not part of the collectively accepted system of rules that is the country, that the status it assigns to itself as trustworthy is true; its signals concerning strategies ought to be trusted.

Another option, making it possible to stick to the institutional definition by Aoki (or North, with some modification of the description below), would be to take one step back and define a game with the extent of a country's ability to signal their commitment to a particular equilibrium strategy in "investor games" as an outcome, with all the different actors inside the country determining this ability as agents. Then, the IQ of the country in the original game defined above would be referring to the self-sustaining system of shared beliefs about how the "what-to-signal-to-outside-investors-game" is played, with higher quality as shared beliefs about sending a stronger signal of respect. Think for example of a common good-game (the common good here being the investor-attracting signal of trustworthiness), where each of K agents (e.g. government, population subset 1,2,3..., business world, etc), can make a contribution in the range of $[0, c]$ toward the common good, in this case making sacrifices through behaving in a trust-inducing way that requires effort. The possible signal strength in G^{∞} is an increasing function of the combined effort from the K agents, or alternatively, the ability to signal at all requires some combined threshold contribution. If for each attracted investor, the signal is less costly, or the combined threshold contribution decreases, the situation is one of games linked with positive complementarities. This seems many times to be the case as much of the effort involved is such that the marginal effort is zero as most of the signaling is not specific for a particular investor, and investors may very well view the presence of other investors in a particular region or country as a signal of less risk, provided that these investors have not suffered any damage from bad institutions.

Looking at the concrete examples, in the finite horizon game above, there is only one possible equilibrium payoff profile and already zero variance of possible payoff profiles; it is not possible to lower the variance of possible payoff profiles through a signal. In this setting, it is simply not possible for the country to send a credible signal about not stealing every period, and the agents are trapped in the unfavorable, unique, equilibrium.

In G^{∞}, there is a large number of possible equilibrias to which the country is, technically, able to signal commitment to. One very important question regards the interpretation of the signal; what it is and how it could be interpreted. But there are unclarities in the technical dimension as well. If there are no restrictions on which equilibrium that the country signals commitment to, if all that might change is the strength of the signal, and any cost associated with the signal is equally low or high regardless of which (set of) equilibrium the signal is aimed at, then it is hard for a rational investor to believe that a rational country would ever signal commitment to any other equilibrium than that in which (*commit*, *steal*) is played its maximum frequency such that the investor still invests. If the country is able to send a signal strong enough to erase all uncertainty about which equilibrium strategy it will use, it will choose to signal the equilibrium that maximizes its payoff. If the cost of signaling varies over different equilibrias, it is no longer possible to say what is optimal for the country to signal, although it is still very likely that there will be one optimal equilibrium for the country to signal commitment to.

Depending on the interpretation of the game, intuition could lead one to assume both higher and lower cost for signaling lower stealing frequency. If for example stealing from time to time is interpreted as taxing the investor, signaling to tax the investor moderately more, moving as far as possible towards the point of negative payoff for the investor, without going beyond it, might be a harder signal to send than to just commit to very low, constant taxes. The investor might reason along the line of 'if the country increases stealing to frequency f, why not $f + 1$ or $f + 2$?' making this signal harder to believe and more complex than setting up a special economic zone where the country promise to tax zero. If stealing on the other hand is interpreted as actual stealing, and the country has had problems with high criminality and occasional looting, signaling to completely eradicate these problems could be harder and more costly (albeit probably giving a higher long-term payoff) than to just mitigate the problems enough to make it worthwhile for the investor to invest, and the logic behind signaling a payoff (stealing) maximizing equilibrium for the country would be reinforced.

Both these lines of reason could be turned around, to instead assume it more costly for a high-taxing country to signaling very low taxes than to signal maintained or moderately increased taxes kept below a specified upper limit, and more costly for a somewhat corrupt country to credibly signal a decrease in criminality and rent-seeking, putting it just above the investor's investing limit, than to take more drastic steps to eliminate these problems and try to signal zero stealing. This would be because it could be harder to believe a high tax country to eliminate taxes than to hold them constant, and harder to believe an essentially corrupt and criminal society to succeed with minor improvements without radically changing the underlying reasons of criminality and rent-seeking, than to believe radical changes altering the underlying reasons to have an effect.

Technically, interpretations in both directions above are possible, and as long as the cost of signaling is not exactly counterbalanced by the varying stealing frequency, there will be one optimal equilibrium for the country to signal commitment to. Another option, where the possibility of many different equilibrias still exist, is if the country is just one of many countries playing this investor-country game with many investors, simultaneously choosing what to signal. Then, there are reasons for a rational country, believing other countries to signal more investor-friendly equilibrias, to also choose a more investor-friendly equilibrium than the one with highest possible stealing frequency, as it otherwise runs the risk of not attracting any investors. This is of course altering the rules of the game, as they were initially stated.

Some suggestions and hints on possible interpretations of signals made by the country are given above. Different interpretations are naturally linked to different interpretations of what the other actions (stealing, respecting, etc.) mean, and if stealing is interpreted as to be anything at all decreasing the payoff for the investor below what it would expect in the respect-case, then the signal could be anything likely to mitigate the risk of this payoff-decreasing event to occur. Improving several of the factors included in ICRG, e.g. better government stability, lower corruption or mitigating religious tensions, could very well be seen as signaling commitment to non-disruptive behavior. Possible connections between ICRG and IQ as credible commitment are discussed further below.

4.3 Institutional Quality as Length of Foresight

In the variations of the game above, we assume that the agents can look into the future to the 'end of time', which might or might not be finite. This is a quite strong assumption, both when it comes to having actual knowledge of the future, and the mental capacity to care as much about future selves as about the current self. Many things associated with IQ of a country could be interpreted as taking account of the future to a greater extent. Examples are work ethics (working harder to get a greater payoff in the future, assuming working hard is not done for the fun of it), trust (as betraying peoples' trust will punish you and society in the future and behaving in a trustworthy way will benefit you in the future), low criminality (taking full account of all the years, including those far away in time, you risk in prison), a functioning legal system (there is most likely savings to be made in the (possibly very) short run by cutting expenditures in the legal system), social insurance institutions (investing in lowering peoples' cost of risky initiatives and future social stability), education system (investing in increased productivity in the future by taking 0 or negative payoff in the present, neglecting any discussion concerning intrinsic value of knowledge), etc. Looking at these examples on an individual level, many dimensions of "succeeding" or being perceived as "successful" involves effort and sacrificing present preference satisfaction for future. Again, obvious examples are studying (educational system) and working hard (work ethics).

Definition 7. IQ of a country is to what extent it has the ability to take the future into account.

This can be done in different game theoretic ways. One (standard) way is to simply introduce a discount factor $0 \leq \delta < 1$, discounting every time period as much as all the other; the value of a payoff of x in period j, t periods into the future from period i is worth exactly δ^t times what the payoff would have been worth in period i, regardless of the values of i,j, and t. Discounting in this way does not qualitatively change the possible equilibrias of the game, assuming δ is sufficiently close to 1[22]. In addition, it is well established that people do not discount in this linear way; neither does supposedly rational investors.[23]

The approach taken here is instead to put a limit on agent's forecasting capabilities. We depart from the game G^{∞}(or G^{fin}) and introduce some new notation:

- A set of possible limited foresight-lengths, $LF = \{lf\}$
- A set of future-functions, containing information about how the agents view the future past their foresight, $F = \{f\}$
- A set of all finite action profiles $\mathbf{A_e}$ of length lf such that an infinite repetition of any $\mathbf{a_e}$ =$[a_e^1, a_e^2, ..., a_e^{lf}]$$\in$ $\mathbf{A_e}$could be supported as an equilibrium in G^{∞}
- Denote by $h^{n:m}$a history with the stealing frequency of n times per m periods.
- Let the modified game be denoted G^{lf}

In this variation of bounded rationality, the agents have limited forecasting capabilities. They are constrained by not having a clear picture of what might

[22]See e.g. theorems 1 and 2 in Fudenberg and Maskin (1986), or Osbourne and Rubinstein (1994), chapter 8, for the Folk Theorems

[23]See e.g. Shane and O'Donoghue (2002), and DellaVigna and Pollet (2007)

happen beyond their horizon of foresight, namely beyond lf periods into the future. At time t, i is able to carry out the standard backward induction thinking lf_i periods into the future. i does not have a clear view of what might happen beyond this point in time, but is aware of that there is (if there is) a future to which i assigns some value. At time t, i assigns value $f_i(h_{t+lf})$ to the unknown future that follows after history h_{t+lf_i},the history of length $t+lf_i$ that i believes to be what will happen as far as i "can see".

Two things are important to notice about the future function:

1. The valuation of the "future" is a function of the history leading up to the future. If the valuation of periods beyond the foresight horizon where independent of what happened up to this horizon, the agents would behave in exactly the same way as in a finite horizon game, taking their horizon as the last period of the game. $f_i(h_{t+lf})$ might depend on h in a large number of ways; the important thing is that the agent believes that his actions affect the future in some way without knowing, or forming an exact opinion about, exactly how.

2. The valuation of the future is bounded and its maximum is increasing with the length of foresight. The intuition behind this is that

   (a) if the agent is not able to grasp infinity time wise, it should not be able to grasp infinite utility either. If it was the case, the possible utility within the horizon of foresight could be comparatively worthless in relation to the possibly infinite utility in the unknown future.

   (b) an agent able to grasp a longer period of time is also able to grasp a greater total utility, and is therefore able to see a higher possible value of the future as well. A more short-sighted agent, in addition to calculating and forming a clear opinion of what will happen fewer periods ahead, also sees less potential value in the future.

We set $f_i(\cdot) \leq max_{\mathbf{a_e} \in \mathbf{A_e}} \sum_{j=t}^{t+lf_i} u_i(a^j)$. The maximum utility of the distant future is no more than what the agent can receive during one "limited horizon-cycle" in the most favorable equilibrium conceivable times a constant, k. Given that this is the maximum "distant future-utility" the agent can take into account, it is also the maximum loss in "distant future-utility" a planned course of actions can inflict upon the agent.

We also set the functions $f_1(h^{n:m})$ to decrease with increasing $\frac{n}{m}$ or higher stealing frequency in the history leading up to the future, and $f_2(h^{n:m})$ to increase with higher stealing frequency, $\frac{n}{m}$, given that the frequency does not go beyond what the investor can accept in equilibrium $(n, m : st_1 \cdot n + r_1 \cdot (m - n) > 0)$. Some arbitrariness of how an agent evaluates the distant future is unavoidable. If one wants to limit the possible future utility at all, it has to be done in some way. Even if the agent would be able to think of a higher utility than our maximum, it a reasonable assumption that the agent would not be able to take into account a greater distant future-utility than it is able to overview and calculate during a time of its limited horizon-length. The argument is not that this is a unique best way of modeling how the agents evaluate the future, but a reasonable way of doing it. An alternative approach is e.g. the one Jehiel (2000) takes, letting the future evaluation be a stochastic function of a "mood parameter" and the action taken in the period in which the evaluation is made.

The limited horizon-foresight also puts some intuitive restrictions on strategies possible for an agent. If an agent has a foresight length of e.g. 5, it would be strange to let the agent formulate strategies explicitly for the next 10 periods, as the foresight length express some kind of limitation of the agent to see and calculate further into the future than this length. Allowing for strategies such as 'steal once every $2 \cdot lf_2$ periods', goes against the purpose of the limited foresight condition, to put a bound on the agents ability to overview, plan, and calculate the actions and their effects further into the future than lf. An agent not calculating and over viewing further into the future than lf will not have a strategy being dependent on a greater number of periods than lf. This is not to say that the agents' strategies does not possibly determine what actions they will take further away than lf. For example the strategy 'steal once every $\frac{lf_2}{2}$ periods' assign an action of the country infinitely far away, but the agent need not realize that there are more than $\frac{lf_2}{2}$ periods to be able to formulate it. On the other hand, formulating a strategy as 'steal once every $2 \cdot lf_2$ periods presuppose an agent clearly grasping time beyond lf, explicitly assigning an action to the next $2 \cdot lf_2$.

Formally, agent $i \in N$ with limited horizon-foresight lf_i is restricted to strategies that could be implemented by a machine with lf_i states, or a machine with complexity lf_i.

Both agents at any time t choose strategies, s_i, as to maximize

$$\sum_{t=k}^{k+lf_i} u_i(a^{t|s}) + f_i(h_{t+lf_i}|s) \tag{1}$$

As the agents cannot overview the entire time that the game might be played, this could result in situations where there is no single period in which any agent would want to change its strategy, but still, with the capability to overview the results of this game over multiple of horizons of foresights, the agent would want to deviate from the chosen course of actions. That is, some kind of self-enforcing non-Nash equilibrium state could come about. Therefore, the equilibrium concepts will be modified slightly.

Definition 8. A Limited-horizon modified Nash Equilibrium (LHNE) in the game, G, is defined by strategies s^*, inducing action profiles a^{s^*} such that at any point in time $k \in T$,

$$\sum_{t=k}^{k+lf_i} u_i(a^{t|s_i^*,s_{-i}^*})+f_i(h_{k+lf_i}|s_i^*,s_{-i}^*) \geq \sum_{t=k}^{k+lf_i} u_i(a^{t|s_i,s_{-i}^*})+f_i(h_{k+lf_i}|s_i,s_{-i}^*) \forall s_i \in S_i, s_i \neq s_i^*, \forall i \in N.$$

Definition 9. A Limited-horizon Modified Subgame Perfect Equilibrium (LH-SPE) of the game G is characterized by strategies s^*, inducing action profiles $a^{s^*}|h$ such that, at any given point in time $k \in T$, and after any history $h \in H$, reached or not

$$\sum_{t=k}^{k+lf_i} u_i(a^{t|s_i^*,s_{-i}^*}|h)+f_i(h_{k+lf_i}|s_i^*,s_{-i}^*,h) \geq \sum_{t=k}^{k+lf_i} u_i(a^{t|s_i,s_{-i}^*}|h)+f_i(h_{t+lf_i}|s_i,s_{-i}^*,h) \ \forall\, s_i \in S_i,\ s_i \neq s_i^*,\ \forall\ i \in N.$$

We start out with the assumption that both agents have correct beliefs of what actions the other player will take on the equilibrium path $min_{i \in N}(lf_i)$ periods into the future. In the case of $lf_i > lf_j$, i knows lf_j and j knows that $lf_i \geq lf_j$

In this game too, there is a wide array of possible equilibrias. Before putting any restrictions on f and lf, the set of possible equilibrias is even larger. This is quite natural considering that we introduce more parameters, and the old situation is a special case of the new, (when $lf \to \infty$, these definitions are equivalent with the definitions above. To see this, note that the $\sum -$ term will cover the entire game, in itself. The $f-$term will be worthless, as there is no future beyond an infinite horizon of foresight; there are no periods not already included in the $\sum -$term).

An equilibrium where the investor invests does have some necessary characteristics:

- At time zero, the investor must judge that investing is worth the investment cost, i.e. $\sum_{t=1}^{lf_1} u_1(a^{t|s^*}) + f_1(h_{lf_1}|s^*) > I$
- At every other point in time, $t \in T$, $t > 0$, the investor will *commit* to the investment if he/she believes that expression (1) is greater than zero. If this condition is not fulfilled, the investor will *leave.*
- The country will *steal* every period in which it does not believe doing so will make the investor leave, i.e. every time it does not think that *stealing* will lead the investor to evaluate (1) as less than zero in that or any coming period. This could be stealing frequencies ranging from every period, if the investor does not even invest, or zero periods of stealing, if deviating from the respect-commit path during one single period would lead the investor to leave.

Trying to characterize the equilibrias in this way is more or less restating the equilibrium definitions with other words. However, interesting situations emerge with some given (f, lf), where the set of possible equilibrias changes significantly. Consider for example the situation when $lf_1 = lf_2 = lf$ and

1. $-st_1 > (lf - 1) \cdot r_1 + f_1(h^{1:lf})$, not one single period of stealing during a limited horizon-cycle can be sustained in equilibrium, which also implies that $f_i(\cdot) \leq r \cdot lf_i$
2. $st_2 > 2 \cdot lf \cdot r_2 \geq lf \cdot r_2 + f_2(h)$, $\forall h \in H$, the second inequality due to the fact that $max\{f_2(\cdot)\} = lf_2 \cdot r_2$. which in turn follows from (1), that no stealing can be sustained in equilibrium in an infinite repetition of a limited horizon-cycle strategy.

In this situation, the country would like to commit to not stealing at all, since the investor will not invest otherwise, but is unable to do so. From the perspective of time 0, both agents are aware of that the country will steal a first time no later than in period lf. Knowing this, the investor will not invest. The insight here is that a too short-sighted agent rules out any possibility of cooperative behavior. Holding everything else constant and increasing lf, i.e. increasing the IQ, we will reach a situation where the second condition is no longer fulfilled and the country respecting the investor is once again possible.

Another situation not possible in the standard setting is the investor taking on the investment cost without ever getting it back. If e.g. the strategy of the country is to steal once every lf_1 periods and respects in all other periods, and $st_1 + (lf_1 - 1) \cdot r_1 + f_1(h^{1:lf_1}|s) > I$, but $st_1 + (lf_1 - 1) \cdot r_1 < 0$. Here, the investor will *invest* the country but loose every limited foresight-cycle, as it always believes that it will start getting positive returns in the future but never does. One objection to this scenario is that it is simply a form of false believes from the investor. In the standard setting, just as well as assuming that $f_1(h^{1:lf_1}|s) > 0$ (which implies that the investor judges the country's stealing frequency to be lower than $\frac{1}{lf_1}$ when in fact it is not), one could in the standard setting assume that the investor holds false beliefs about the strategy utilized by the country. If f_1 was explicitly defined, this would however be a more elaborate way of designing the nature of the agents bounded rationality, following some kind of pattern, instead of just saying that an agent holds false beliefs about the future. Rather than seeing this as an argument for abandoning putting a limit on agents' foresights, this can be viewed as an argument for restricting the way in which the agent evaluates the unknown future, making explicit the form of f.

Definition 10. Relative IQ of the country C is defined as the length of lf_C , or how distant future the agents are able to take into account. The IQ of C is higher than that of C' if and only if $lf_C > lf_{C'}$.

The main take away from this setup is that with a too short time horizon paired with significant gains from stealing, cooperative behavior could be ruled out, much like in the finite horizon game. The difference here is that the agents are aware of that there is a next period, and a next period after that, they just lack the ability to take it into account. There are numerous possible real life situations like this, with an agent knowing a period not to be the last one but failing to take it into account to the extent it ought to. One is listed companies failing to look as far beyond the quarterly report-logic as they should, shown e.g. by DellaVigna and Pollet in "Demographics and Industry Returns", where by using publicly available demographic data, investors could achieve above industry returns by investing in school buses more than five years before unusually large year cohorts reach school age but fail to do so. Another might be jail sentences losing their additional deterring effect beyond a certain threshold when criminals fail to take the additional time in jail into account. They know for a fact that there is a time beyond e.g. 7 years (assuming they will live that long), but might still fail to take into account the additional punishment the last 3 years of a 10-year sentence constitute relative to a 7-year sentence; this due to their limited horizon. A third example could be short-sighted politicians failing to take account of long-term effects; two Swedish examples are the 'Million Program', and the economic policies used during the 70's and 80's with inflationary policies, devaluation, subsidizing borrowing, and unpredictable changes in general. A possible objection to the third example is that, rather than short-sightedness of politicians, it is a manifestation of self-interested politicians only caring about winning the next election. An answer to this is to redefine the statement to say that it is the democratic institution or political system, rather than individual politicians, that is manifesting the short-sightedness.

It is possible with a scenario where the investor utility-wise benefits from a country with decreased time horizon. In the case of an investor able to accept stealing once every lf_1periods, and a country with time horizon $lf_2 = lf_1$, an equilibrium where an investment is made and the country respects $lf_2 - 1$out of lf_2 periods is possible. If lf_2 is decreased to $lf_2^{'} < lf_2$, so that a stealing frequency of once every $lf_2^{'}$cannot be accepted by the investor, but the time horizon is not so short, proportionally to the payoff from stealing $(st_2 + (lf_2^{'} - 1) \cdot r_2 + f_2(h^{1:lf_2^{'}}) < lf_2^{'} \cdot r_2 + f_2(h^{0:lf_2^{'}}))$, so that the country necessarily end up stealing. This would rule out other equilibrias except the investing with no stealing at all, and the investor not engaging with the country which steals every period.

An interesting note concerned to this setup is that the investor actually could benefit from being more short-sighted as this could rule out any stealing at all. Interpreting stealing as e.g. taxing investors, an interpretation of this could be that it is what happens when countries set up special economic zones where foreign investors are exempted from taxes: very short-sighted investors acting in accordance with the rules of quarterly capitalism will not be able to accept any "stealing" at all in order to invest, thereby effectively forcing for the countries less than optimal conditions upon them through the short-sightedness.

Even though, by definition, utility maximization, and therefore utility-maximizing institutions, is what is desired, looking at IQ as the length of time horizon in a particular setting could help in focus on key things to induce a cooperative (or utility maximizing) behavior. Albeit this is a detour from just defining IQ as utility maximizing institutions, as we have seen above, this does usually not provide any insight in what might be prohibiting, or catalyzing, utility maximizing behavior, or how a simple and reasonable restriction as a limited horizon of foresight may alter results.

This definition refers to an actual agent of the game and intuitively needs a definition of institutions as the one Searle gives. Interpreting it as if it is a very specific subset of the 'rules of the game', namely the foresight length, one could possibly use North's definition of institutions, although it is quite doubtful, both because he makes a point specifically out of the separation of organizations and the rules they operate under, and because one specific rule of the game hardly is 'the rules of the game'.

IQ as Length of Horizon of Foresight With Actual Length Unknown
Introducing communication in a non-cooperative game theoretic setting is not unproblematic. Exactly how the country, in the setting of *4.2,* is able to credibly signal commitment is unclear. Here, we will instead look at a signaling possibility of the country similar to "burning money", with the disparity from the game in which it was first proposed that here with limited and unknown length of foresight, it could be relevant for the situation. Consider the same situation as G^{lf} with the difference that lf is not known. We introduce some additional notation:

- A set of states, Ω. Each $\omega \in \Omega$ is associated with a unique combination of $\{lf_1, lf_2, f_1, f_2\}$
- A set of partitional information functions, $P = \{P_1, P_2\}$, of which each associates a non-empty subset $P_i(\omega) \subseteq \Omega$with every $\omega \in \Omega$. We also assume common knowledge of P.
- A new "point in time", -1 is added, so that $T = \{-1, 0, 1, ...\}$
  - A new action *burn* is introduced for the country, $A_2 = \{Burn, Respect, Steal\}$ where *Burn* is available at $t = -1$ and *Respect* and *Steal at all times* $t > 0$.

Consider a situation with two states, ω^{long} and ω^{short} associated with $\{lf_1, lf_2^{long}, f_1, f_2\}$ and $\{lf_1, lf_2^{short}, f_1, f_2\}$ respectively, where $lf_2^{long} > lf_2^{short}$ and for simplicity, also $lf_1 > lf_2^{short}$. Assume that if we are in the state ω^{short}, we are in the situation from above, where

$$-st_1 > (lf_1 - 1) \cdot r_1 + f_1(h^{1:lf_1})$$

(not one single period of stealing during a horizon-cycle can be sustained in equilibrium) and a too short-sighted country keeps the investor from investing.

$$st_2 > 2 \cdot lf_2^{short} \cdot r_2$$

In the state ω^{long}, on the other hand:

$$st_2 < 2 \cdot lf_2^{long} \cdot r_2 \vee st_2 < lf_2^{long} \cdot r_2 + f_2(h^{0:lf_2^{long}})$$

The information partition is such that $P_2(\omega^{long}) = \omega^{long}$, $P_2(\omega^{short}) = \omega^{short}$, $P_1(\omega^{short}) = P_1(\omega^{long}) = \{\omega^{long}, \omega^{short}\}$; there is no uncertainty for the country concerning which state of the world it is in, but the investor does not know whether the country is short-sighted or not before investing (once the investor has invested, it can infer which state it is from the strategies used). We also assume common knowledge of P: the investor knows that the country knows if it is short-sighted or not, and the country knows that the investor does not. The IQ of the country, or investor-country game, is unclear. This situation could very well discourage the investor from investing, not wanting to risk investing in a short-sighted country that will likely end up stealing. Adding the option of "burning money", giving the country a payoff of $-b$ in period -1 could change the situation:

1. When the state is ω^{short}, as

   $$st_2 > 2 \cdot lf_2^{short} \cdot r_2 = lf_2^{short} \cdot r_2 + max(f_2(\cdot))$$

   any equilibrium strategy for the country must include at least one period of stealing, as deviating would increase (1). At the same time, since

   $$-st_1 > (lf_1 - 1) \cdot r_1 + f_1(h^{1:lf_1})$$

the investor will increase (1) by leaving prior to the stealing. I.e. in state ω^{short}, if the investor happens to invest not knowing which state it is in, it will leave instantly and the country will steal instantly.

2. If the state is instead ω^{long}, and

$$st_2 + (lf_2^{long} - 1) \cdot r_2 + f_2(h^{1:lf_2^{long}}) < lf_2^{long} \cdot r_2 + f_2(h^{0:lf_2^{long}})$$

strategies where $(respect, commit)$ always is played could be sustained in equilibrium.

An investor not knowing which horizon of foresight the country has might or might not invest, depending on degree of risk aversion and probabilities attached to the respective states. E.g. a risk neutral investor believing

$$p(\omega^{long}) \cdot (lf_1 \cdot r_1 + f_1(h^{0:lf_1})) + p(\omega^{short}) \cdot st_1 > -I$$

where $p(\omega)$ denotes the probability the investor assigns to the state ω, would invest in spite of the uncertainty.

Looking instead on the case when

$$p(\omega^{long}) \cdot (lf_1 \cdot r_1 + f_1(h^{0:lf_1})) + p(\omega^{short}) \cdot st_1 < -I$$

the investor will not invest. However, if we add the primitive signaling option of burning money for the country, the situation may change. Assume

$$st_2 < b < lf_2^{long} \cdot r_2 + f_2(h^{0:lf_2^{long}})$$

that is, a country with horizon foresight lf^{short} could never expect to make up for the negative payoff but a country with horizon foresight lf^{long} could. By simply inflicting damage upon itself, it communicates that it has a horizon of foresight long enough to make up for the damage, thereby ruling out the forced uncooperative behavior of ω^{short}, resulting in the investor investing.

Often in the literature, the favorability of stable institutions is emphasized. Assuming that people, technology and environment changes, the optimal institutions adjusted for a specific set of people technology and environment, should also change. Keeping these constant would be bad or costly, rather than a good thing. One interpretation of static or stable institutions is as a signal of non-short-sightedness on the side of the country, that helps agents' coordinate their actions, as well as acting non-short-sighted themselves. Maintaining stable institutions, however, is not actually hurting anyone in the direct way that could naturally be interpreted as 'burning money'. Other candidates to interpret as analog to burning money could be countries siding against part of their populations, moving villages to create special economic zones, or make way for specific investments. These come closer to possibly be interpreted as actually inflicting hurt upon the country in the short run to out rule any suspicion of being too short-sighted.

4.4 Compatibility and Complementarity of the Definitions

The discussion below is mainly about compatibility in the technical sense. Looking at the definitions more from a philosophical, or non-technical intuitive way, there is not really anything about IQ according to one definition that with necessity makes IQ according to another wrong or completely superfluous. One might very well consider the different definitions of IQ as utility, credible commitment, and length of foresight as different parts of IQ that are all interesting to an investor contemplating to invest in a country. An investor will likely care about possible payoffs (utility), how uncertain the payoffs seem today (credible commitment) and the risk of abrupt changes in future behavior (credible commitment and length of foresight, as the risk of a short-sighted country to make abrupt changes and stealing is larger in general than that of a far-sighted country). In this sense, they are all complementary and each have their own place in discussing IQ of a country.

4.4.1 Utility Maximization and Credible Commitment

A first question concerning whether the different definitions are compatible or not, is if the definitions of institutions that they are based on are compatible. IQ based on different definitions of institutions can hardly be consistent with each other in a strict sense, as the entity to which the quality refers is in itself different in those cases. An investor looking at a situation cannot at the same time consider mean utility of possible equilibrias on the one hand and the country's ability to signal commitment to a certain strategy on the other, to define IQ. Even a view of them as different parts of IQ is highly problematic as the institution possessing quality of a certain level is different. In the first case, the institution is the rules of the game, in the second, the country itself.

Another basic reason for the definitions to be incompatible is that they may very well end up with different orderings of where the better IQ is, and disagree on what is improving or deteriorating IQ. E.g. a country with the ability to eliminate all uncertainty by its signal σ credibly committing to a strategy s^*, could decrease the mean of possible equilibrium payoffs, i.e. a higher or improved IQ according to definition 5 would in this case correspond to lower or worsened IQ according to definition 3 and 4. This said, for certain specifications of parameters, the definitions are compatible in the sense that an improvement in IQ according to one of the definitions is so according to the other as well.

When defining IQ as shared beliefs about playing the game in a way maximizing the utility for the investor, there is no longer any uncertainty concerning what will be played. A signal like the one in definition 6 could be (part of) the reason why a certain set of shared beliefs converge toward playing the game in a specific way. However, definition 4 presupposes that the beliefs will converge and is in no need of any signal to lower uncertainty that no longer exist once the institution the definition refers to, the shared beliefs, has been formed.

4.4.2 Utility Maximization and Length of Foresight

The same kind of objection to compatibility between length of foresight definitions and utility definitions (both definition three and four) of IQ apply: they build on definitions of institutions that are separate, and there are situations with combinations of parameters where an ordering or improvement in IQ according to one definition is a different ordering or worsening according to the other. For other parameter specifications, the definitions may also coincide in changes and ordering of institutional quality.

4.4.3 Credible Commitment and Length of Foresight

Definitions of IQ as signaling credible commitment and length of foresight, on the other hand, build on the same kind of definition of institutions, and capture different characteristics of the institution they refer to. In the technical world the definitions are given, the signal a country is able to send does not affect its length of foresight, neither does the length of foresight directly affect the signal.

The length of foresight affects which equilibrias that are possible, and therefore which strategies that a country is able to signal credible commitment to, but not the strength of the signal. An increase in IQ according to definition 9 only gives new equilibrium possibilities, it does not take away any. Hence, there are only more equilibrium variance to remove with the signal, and greater potential for IQ according to definition 6. Moving in the opposite direction, with decreasing length of foresight and IQ according to definition 9, we could move to a situation where there is only one possible equilibrium left and no room for any credible commitment signaling at all, which does not exactly contradict definition 6, but makes it superfluous. A possible contradiction could result if a country's signaling ability has a very specific structure based on possible equilibrias in a certain setting. For example with equilibrias e^1, e^2, e^3 possible, the country is able to send a signal eliminating all uncertainty, but when lf_C is increased so that also e^4 is possible, the country is no longer able to s end a strong signal. In this specific setting the definitions contradict each other, but with signaling ability independent of available equilibrias, there is no incompatibility in the form of contradicting institutional between the two.

4.5 An Empirical Example: ICRG and its Connection and Compatibility with the Definitions

As noted above, ICRG is an index used to assess the country specific risk ratings aimed at international business such as multinational firms, banks and equity funds. The index is made up of the three overall categories of *Political, Economic,* and *Financial* risk, each made up of 5-12 components, of which some in turn are made up by a few sub-components. The individual components of the economic and financial indicators are intrinsically quantitative, objective measures, to the extent that the data is trustworthy (which they are not always, as e.g. the 2010 debt crisis in Greece has shown). The political ones are subjective and leave more room for interpretation as they may be quantified in many different ways. The comprehensive country risk rating is basically a weighted average of all the individual measures, shown in table below.

Whether or not different measures actually are part of the institutional framework is not completely clear. The financial and economic factors are in general more of outcomes of institutions, or input to institutional development, rather than institutions themselves. GDP per capita for example could be a proxy for

Political	
Government Stability	Military in Politics
Socioeconomic Conditions	Religious Tensions
Investment Profile	Law and Order
Internal Conflict	Ethnic Tensions
External Conflict	Democratic Accountability
Corruption	Bureaucracy Quality

Table 3: Political Components of the International Country Risk Guide (Table 1 repeated)

Economic	Financial
GDP per Capita	Foreign Debt, % of GDP
Real Annual GDP Growth	Foreign Debt Service, % of Exports
Annual Inflation Rate	Current Account, % of Exports
Budget Balance, % of GDP	Net International Liquidity, Months of Import Cover
Current Account, % of GDP	Exchange Rate Stability

Table 4: Economic and Financial Components of the International Country Risk Guide (Table 2 repeated)

good IQ, and might very well be a cause or catalyst of IQ improvements (see e.g. Glaeser et al 2004, Chong and Caldéron 2000 for discussions concerning increased GDP as cause of improved IQ), but is not in itself an institution or a building block of an institution if one use any one of the three definitions of institutions presented above.

The political ones on the other hand are in general factors such that they could be described as part of the rules of the game, or part of a self sustaining system about how the game is played. For e.g. an investor contemplating whether or not to invest in a department store, whether or not civil unrest and looting is a significant risk or not, is highly relevant for the rules of the game, or the self-sustaining system about how the game is played. The response to this may be to invest heavily in security guards, or not to invest at all; in both cases it affects how the investor 'plays'. Exactly which part of institutions and IQ that investors care the most about is likely to vary with the kind of investment, investor and country we are concerned with.

Utility Maximizing Institutions The correspondence between good ICRG-ratings of a country, and choosing, for an outside investor, a payoff maximizing course of action is not very clear; the definition is general to the extent where it is hard to say too much about it. The countries with highest ICRG-ratings are rich democratic states with developed economies. These are not in general the ones providing investment opportunities with the lowest stealing frequency, if one includes taxing as a mild form of stealing. It is however unlikely that these countries would exhibit the more severe forms of stealing such as actual looting; low ICRG-ratings will most likely be the countries associated with both the lowest and highest institutional quality in the utility-maximizing sense in the game described above. Precisely because these are the more risky alternatives, they need to have some opportunities for very high returns, as the risk of getting negative returns is greater here than in the more stable developed states. The high ICRG institutional quality countries will probably be the mid-range countries associated with some stealing (high-taxing), but with considerably less variation. Including taxes in the stealing part, ICRG-quality and utility maximizing-quality do not fit well together.

If one however sees low taxes and high possible returns as part of r_1 instead, the likely correspondence between ICRG and utility maximization institutional quality is considerably closer. This would probably be reinforced if the utility of the country in the investor-country game was included in the utility-maximization part, as the high ICRG rated countries either have less trouble with attracting capital or are (generally) in less dire need of capital, and are therefore in a better position to put their own interests first.

Credible Commitment Given that indexes such as ICRG actually is used by investors to assess risk in countries, actions by a country to improve its ICRG-measure is in itself a signal of committing to a non-value-destroying behavior by the country, regardless of whether the signal is true or not. If investors actually use and trust measures such as ICRG, improving this measure is a signal of credible commitment, even if it is not actually credible commitment.

Most likely though, it is a true signal about decreased risk and commitment to not choosing a too disruptive behavior. Naturally, the individual indicators that make up the index are chosen to represent qualities making countries and thereby investments in these countries less risky. Unexpected expropriation is less likely with a stable government with democratic accountability; law and order, and socioeconomic conditions decrease the risk of looting and general crime; conflicts and ethnic and religious tension increase the risk of civil war which would severely hurt basically any investment (unless one is actually investing in war, through e.g. arms or mercenaries); budget balance, current account balance and low inflation rates decrease the risk of profits being stolen by inflation and a deteriorating exchange rate (if the investment is such that the profits should come from inside the country); etc. What the investor really needs to know to invest is not the exact course of action the country will take, it is to know that the country's stealing frequency is low enough so that the payoff from investing remains positive. This is a considerably less complex thing to signal, and corresponds well to various indexes, such as ICRG.

Length of Foresight Just as IQ as signaling credible commitment, IQ as length of foresight goes well together with a practical application of IQ as ICRG. Many of the political indicators could be signals of non-short-sightedness in themselves, or enablers of non-short-sightedness, e.g. government stability, internal conflict, and law and order. Occasionally, it could even be that improving some factor, e.g. investment profile (itself made up of contract viability/expropriation, profits repatriation, and payment delays), was made in part to actually signal a longer length of foresight and attract capital. In the very short term, many countries could probably benefit from expropriating assets of many foreign investors; looking further ahead than the very short term, it would most likely be harmful.

Things need not be intended to be a signal of a longer time horizon to actually function as one. Improvements of most of the different factors of the ICRG seem somewhat far-fetched to interpret as actual direct attempts to signal a longer time horizon to outside investors. For example democratic accountability - how responsive government is to its population - may be improved by a state for a multitude of reasons, like the people demanding it, government sincerely caring about the peoples' opinions, belief in the moral principles behind it, etc. That signaling non-short-sightedness to external investors is the actual main underlying reason appears unlikely. Nevertheless, it may signal this, as democratically accountable states are less likely to suffer coup d'états or other highly disruptive events. The same goes for many of the other factors in the ICRG index not changed with the direct purpose of signaling to outside investors.

4.6 Generalization of the Definitions

The definitions of IQ have, for the most part, quite obvious generalizations in a technical sense. It is easy to re-define them to a general setting. The problem is that somewhat strange and unintuitive interpretations arise, and it might be quite hard to compare IQ from one setting to another without running into trouble, such as assuming cardinal utility instead of ordinal. The definitions below are suggestions of generalizations. They are discussed only to a very limited extent.

Definition 11. *Corresponding to definition 3.* With institutions as 'the rules of the game', IQ is defined as to what extent the rules of the game in a particular situation, provides a set of equilibrium opportunities with the highest possible arithmetic mean of the total utility derived by the agents in the different equilibrias.

Definition 12. *Corresponding to definition 4.* With institutions as 'a self-sustaining system of shared beliefs about how the game is played', IQ is defined as to what extent the shared beliefs converge towards that the game is played in a utility-maximizing way.

The generalization of definitions 5 and 6 presupposes the agent i to be an agent with the capability to create institutional facts, in the sense Searle talks about. Otherwise, the definition does not apply.

Definition 13. *Corresponding to definition 5.* IQ is to what extent an agent i is able to send a credible signal to other agents, N_{-i} about which equilibrium strategy, s_i^*, i will use.

Definition 14. *Corresponding to definition 6.* IQ of an agent is to what extent it has the ability to take the future into account.

Definition 15. *Corresponding to definition 9.* Relative IQ in the game G^{lf} is defined as the length of $lf_{min} = min_{i \in N}(lf_1, lf_2, ..., lf_n)$, or how distant future the agents are able to take into account. The IQ in G^{lf} is higher than in $G^{lf'}$ if and only if $lf_{min} > lf^{'}_{min}$.

An example of problems arising with a generalized definition is for example in definition 11 and 12, where the actual values of the utility of an agent becomes relevant to order institutional quality. In contrast, e.g. with definition 3, it is enough to order outcomes of one agent to determine which of two situations that provide greater institutional quality, and the numerical values are hence not relevant for the ordering of institutional quality, only the utility ordering.

5 Discussion

This has been an attempt to introduce game theoretic definitions of IQ and analyze the respective models. The main part of the discussion and analysis is given in the respective section to which it belongs. Here, a short overview of the discussion is given.

It is hard to conclude too much concrete and specific from the analysis, such as pointing out one definition as correct, or better than the others. The models each have their merits and drawbacks. The utility maximization definition is basically true by construction and a natural starting point, but provide little insight, mainly precisely because it is true by construction. The other definitions on the other hand capture less obvious dimensions and mechanisms that investors care about when looking at a country, but may also provide a false image if in a particular case an investor actually care much more about some other institutional dimension than credible commitment or length of foresight.

The main conclusion is basically reiterating and emphasizing the importance of analyzing each individual scenario to try and find out what the different agents perceive as the relevant factors of their economic reality and try to give a formal game theoretic description of the situation that fits this perception. The models given above with IQ as credible commitment, or burning money to demonstrate non-short-sightedness are possible examples of models that could explain certain behavior from developing economies towards their own population, such as land grabbing. They deem attracting investors important enough to actually hurt their own population in the short run (there could of course be different explanations for behavior like this, e.g. corrupt governments seeking profits for themselves).

An obvious extension of this work, unfortunately outside of the scope of this paper, is to try to find data over time on proxies for credible commitment and varying lengths of foresight, combined with data on investments, and match this empirical data with models of IQ to see how they measure up against reality. Exactly what these proxies could be are not clear; different combinations of ICRG components is one possible starting point.

Another extension could be to develop a compelling model of IQ in linked games, as individual games are rarely able to capture all economically relevant factors of a certain situation.

Facilitating a discussion concerning what is meant by IQ is an end in itself. Many times disagreements have deeper roots than the setting in which they are being discussed. Actually trying to agree on definitions of institutions and IQ could improve discussions concerning institutions by leading to the core of conflicts, thereby helping to avoid fruitless confusions rooted in deeper (and many times less controversial) disagreements.

6 References

Aoki, Masahiko, (2001), "*Toward a Comparative Institutional Analysis*", MIT Press Cambridge, MA, USA

Binmore, Ken, (2010), "Game Theory and Institutions", *Journal of Comparative Economics,* 38 (3), 245-252

Calderón, César; Chong, Alberto, (2000), "Causality and Feedback Between Institutional Measures and Economic Growth", *Economics and Politics,* 12 (1), 69-81

DellaVigna, Stefano; Pollet, Joshua, (2007) "Demographics and Industry Returns", *The American Economic Review,* 97 (5), 1667-1702

Fudenberg, Drew; Maskin, Eric, (1986), "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information", *Econometrica* 54 (3), 533-554

Glaeser, Edward; La Porta, Rafael; Lopez-de-Silanes, Florencio; Schleifer, Andrei, (2004), "Do Institutions Cause Growth?", *Journal of Economic Growth,* 9 (3), 271-303

Jehiel, Philippe, (2001), "Limited Foresight May Force Cooperation", *Review of Economic Studies,* 68, 369-391

Levchenko, Andrei, (2007), "Institutional Quality and International Trade", *Review of Economic Studies,* 74, 791-819

North, Douglas, (1990), "*Institutions, Institutional Change, and Economic Performance",* Cambridge University Press, Cambridge, UK

Osbourne, Martin; Rubinstein, Ariel, (1994), *"A Course in Game Theory"*, MIT Press, Cambridge, MA, USA

Rodrik, Dani, (2007), "*One Economics, Many Recipes: Globalization, institutions and economic growth",* Princeton University Press, Princeton, NJ, USA

Searle, John, (2005), "What is an institution?", *Journal of Institutional Economics,* 1, 1-22

Shane, Frederick; O'Donoghue, Ted, (2002) "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature*, 40 (2), 351-401

Wikipedia, "Institutional Economics", http://en.wikipedia.org/wiki/Institutional_economics (2011-09-17)

Wikipedia, "New Institutional Economics", http://en.wikipedia.org/wiki/New_institutional_economics (2011-09-17)