Norms and the Dynamics of Institution Formation*

Patrick Francois

Department of Economics,
University of British Columbia and CentER, Tilburg University
and
CIAR, CEPR, CMPO

December 2006

Abstract

This paper analyzes the dynamic interaction between norms – internalized restraints on opportunistic behavior – and institutions – restraints on such behavior deriving from external enforcement. When individuals following a norm suffer pecuniary losses to doing so, the norm is eroded. Institutions, on the other hand, are strengthened when institution designers are rewarded for improving them. The dynamic interaction between these two factors leads to both good steady states with functioning institutions, widespread norm compliance, and trade, and bad steady states where trade breaks down, institutions are dysfunctional and beneficial norms are violated. The model here shows the situations that lead economies to converge on good steady states rather than bad ones; why countries with a history of institutional success are more likely to be successful in future, why countries with a history of failure will require better institutions to acheive even the same level of compliance, and why functional institutions may not be readily transplated from successful to unsuccessful countries.

*I thank seminar participants at British Columbia, Simon Fraser BMO series, Calgary, Dalhousie, Laval, Northwestern, West-Coast Macro Conference, PPP conference Bristol, Public Finance Conference London School of Economics, CIAR IOG Vancouver meetings. I am especially grateful to Siwan Anderson, Tim Besley, Maitreesh Ghatak, Paul Grout, Ashok Kotwal, Kiminori Matsuyama, Nathan Nunn, Lars Osberg and Andrea Prat for helpful discussions.

1 Introduction

This paper studies institutions that function to facilitate and oversee trade. It is concerned with how such institutions are formed when new trading opportunities open up. An illustration of the issues that are the present paper's concern is provided by developments in rural Maharashtra, India, over the last twenty years. Though the home state of Mumbai, India's largest city and financial capital, Maharashtra still has a largely rural population of over 100 million people. Over the last twenty years, it has experienced rapid and ongoing increases in agricultural productivity. Per capita annual growth rates in excess of 5% have lead to large changes in agricultural wealth, and to a massive decline in rural poverty. However, this growth has not occurred uniformly across villages in the state. At the risk of oversimplification, villages seem to have fallen into one of two categories. On the one hand, there are those villages where things have gone well. Many of the technological and trading opportunities that have arrived over the last twenty years have been seized upon. Agricultural productivity is high, farmers produce cash crops that are often sold on world markets, and labor shortages lead to in-migration of seasonal labor from all over the state. On the other hand there are those villages that have hardly changed at all. Farmers in them have persisted with the production of low productivity subsistence crops. Trade outside the village is minimal, and landless laborers experience many days annually of underemployment.

The successful villages are thick with market trade and, by and large, traders expect markets to work. When new trading opportunities arise, for example the recent chance to sell table-grapes in Europe's off-season, entrepreneurial farmers have entered to exploit potential gains. Though filled with numerous glitches at inception, the high initital returns continue to lure entrants. The glitches arise because the institutions required to facilitate, regulate and oversee the new trades do not initially exist. But since traders in these villages generally trust the good-faith intentions of their trading partners, they are willing to risk production nonetheless. These early trades do not always work out well, but they do allow the institutional shortcomings to be exposed, and the surplus from them creates a group of constituents who demand the institutional improvements that will make the trade work better. A virtuous circle is established. High initial levels of trust beget trade, trade creates surplus, exposes problems and leads to demands for institutional improvement, improvements increase the volume of trade, and because institutions eventually start to work, the trust in trading partners is reinforced and continues.

In stark contrast, the villages that have been stagnant are characterized by endemic mistrust. Noone will enter into a new trading opportunity fearing that some of the numerous opportunities for hold-up will be exploited by trading partners. Since noone trades in these new markets, there is virtually no political or economic demand for institutional improvement. However, without institutional improvement, there is no possibility of the oversight required to make the new trade work. Consequently, villagers rationally continue to mistrust trading partners, and rationally continue to use traditional, low productivity, methods that leave

them safe from exploitation but poor.

The key ingredients here are new technological opportunities, norms of behavior that encourage trust, and institutional development. At present, there is no model in economics which can help us understand how these ingredients come together in some places to make development work, whereas in others they fail. The present paper's first aim is to provide a model to do that. The model is not, first and foremost, predictive, but is instead aimed to help us think simply, and with some rigour, about observations like those above.

Central elements of this example drawn from rural India are widespread and play out in observed institutional development elsewhere. I will present other examples featuring the basic interplay between new technologies, norms that allow trust, and ensuing trade that engenders institutional development. The common elements from these examples, which are extremely diverse in origin, suggest that the concerns of focus here may be ubiquitous, and hence that such modeling may offer up broader lessons. The paper's second aim is to explore these. I seek to understand what factors allow some places to embark on the virtuous self-reinforcing path to institutional development and high productivity, and what leads others towards economic and institutional stagnation.

The terms "norms" and "institutions" are central to the analysis here, but generally vague concepts which I wish to avoid conflating.[1] I use them as follows: *norms* are internalized restraints on opportunistic behavior; *institutions* are the restraints on such behavior deriving from external enforcement.[2] The analysis rests on two key assumptions about how they change. (1) Economic incentives enculturate individuals towards norm following behavior. Specifically, when rewards are persistently higher (lower) for those who cheat, then cheating (honesty) becomes a norm of behavior. It can and will change, when incentives change, but such changes take time. (2) When new trading opportunities arise, the new institutions required to oversee the trade also take time to develop. Moreover, the individuals whose efforts determine institutional development are rewarded out of the surplus that such trade generates. When the surplus generated by trade is high (low), rewards are high (low), and institutions develop more quickly (slowly).

The following examples share key ingredients with contemporary Maharashtra:

- In the early 2000's a wave of corporate scandals hit the US. Through questionable (but generally legal) accounting, and manipulation of earnings statements, managers of firms in new economy areas increased the short term valuation of companies to either exercise personal stock options or finance acquisitions. The organizations were mainly internet and IT related, where the newness of the product,

[1] Institutions, according to North (1990), are the rules of the game, and they comprise formal laws, informal constraints, shared norms, beliefs and self-imposed limits on behavior and their enforcement characteristics; i.e., courts, police, judgemental aunts, etc. This broad definition bundles norms together with institutions and is also favored by Greif (2006) "An institution is a system of rules, beliefs, norms and organizations that together generate a regularity of social behavior." p.30. Though I have no argument with this treatment in general, for my purposes here it will be vital to delineate between them.

[2] I am focused on the development of those institutions that underpin impersonal exchange by third-party enforcement. These are of paramount importance since the widespread nature of this exchange in the West contrasts significantly with exchange in the less developed world. According to North (1990, p.35), who also focuses on explaining these institutions, this exchange has been "The critical underpinning of successful modern economies..."

and its unfamiliarity, made valuations and financial flows difficult to assess: for example, Enron, Global Crossing, Tyco, Qwest, Adelphia (see Vranceanau 2003). By all accounts, this period saw a massive surge in wrong-doing. The corporate Fraud Task-Force set up in 2002 within the DOJ charged 354 defendants with corporate fraud in 169 cases and obtained over 250 convictions or guilty pleas within a year. The SEC filed almost 50% more financial fraud cases in 2003 than in the previous fiscal year.[3] The very public increase in cases lead to a decline in shareholder confidence. A Pew Research Center opinion poll conducted through the Harvard Business School found that only 23% of Americans trusted bosses of large companies, in contrast with 75% trusting bosses of smaller ones.

- In the late 80's and early 90's, globalized supply chains allowed outsourcing large parts of U.S. apparel manufacture to low wage economies. Compared with stringent domestic labor standards, the lack of off-shore US regulatory oversight lead some firms to source garments from suppliers employing workers in unsafe or exploitative conditions. These transgressions were eventually exposed by the media. The perceived ethical failures on the part of industry leaders were the focus of public campaigns targetted at some of the biggest manufacturers including: Nike, Liz Claiborne, Reebok, Disney, The Gap and the Kathie Lee Gifford clothes line for Wal-Mart.[4]

- Financing of the 1890s US railroad expansion occurred largely through credit contracts. Though offering greater flexibility, equity contracts were shunned because investors feared the possibility of expropriation by unscrupulous traders. A combination of lax accounting standards, insufficient regulatory oversight, and improperly inventoried railway stocks saw a number of investor fraud cases. Re-capitalization of the industry after the first great depression was stymied by a lack of investor confidence in the "ethical probity" of business leaders.[5]

In these examples, a potentially lucrative new trading opportunity exceeds the limits of existing institutional oversight. Weak institutions combined with perceived short falls in ethical behavior serve to limit the scope of trade: Shareholders want to be able to lend money to firms that exploit new economy ventures, but cannot trust the probity of the individuals running those firms, nor the financial statments of their auditors; U.S. apparel consumers of the 1980s want to buy goods sourced from countries with lower labor costs, but cannot trust that the savings do not arise from compromises in worker safety, or exploitation; investors in the 1890 US railroads would have preferred equity financing of railway firms, but instead were limited to debt contracts as they could not sufficiently monitor financial flows, asset holdings, nor the actions of railroad managers.

In each one of these cases, some trade was undertaken because not all individuals were expected to

[3] See Vranceanu (2003) for more detail on these figures.
[4] See Bartley (2003) for further details.
[5] See Chandar and Miranti (2005) for further details.

be ruthlessly opportunistic, but the listed problems in institutional oversight limited its scope. A set of regulatory reactions intended to improve the institutions supporting trade were set in motion by the perceived shortcomings. The dotcom corporate scandals lead to the establishment of federal task forces of investigation and to an act of Congress – Sarbanes-Oxley of 2002 – increasing oversight and punishments for misleading accounting, and reducing the possibilty of firms influencing their own auditors. The apparel out-sourcing scandals lead first to the establishment of voluntary compliance protocols, followed by formal third-party monitoring, and finally the extension of oversight powers to existing labor standards associations. The early 1900s saw a series of federal regulatory and oversight initiatives with four land-mark pieces of legislation aimed at easing investor concerns.[6]

If effective enough, such regulatory and institutional responses can align agents' incentives and allow realization of gains from trade. The industry apparel responses seems to have been successful in overseeing labor standards. The improved accounting and oversight regulations under the Sarbanes-Oxley act have appeased most investor groups. Legislative initiatives precipitated by the 1890s depression eventually (by the late 1920s) lead to the re-capitalization of the US railways using equity contracts.

The fact that these three examples are drawn from the US and seem to be success stories, is somewhat misleading regarding the possibilities of institution formation in general. The US has a history of successful institution formation. In many countries without such a history, things often play out differently. Consider, for example, the Kuna of Eastern Panama, as described by Sandner (2003). The Kuna are a resource based indigenous community deriving much of their livelihood from the sea. Traditionally, resource stocks were maintained by a combination of informal rules and taboos limiting over-harvesting, restricting entry to breeding areas, and curtailing fishing for market trade. As in the US examples, the emergence of a new, and potentially lucrative trading opportunity, began to test the limits of existing institutional restraints. Improved refrigeration and transport of sea products to city dwellers lead to high prices and increased exploitation of the resource. The high pecuniary gains from norm breaking saw the gradual break-down of most conservation norms – which now play almost no role in regulating the resource.

The Kuna's governing body have been keen to exploit trade with Panamanian city dwellers, while guarding against the over-fishing and resource depletion arising from the erosion of conservation norms. They have experimented with Western style prohibition methods to protect diminishing stocks: Pigovian taxes, increased monitoring, prohibitions of some harvest methods, and closed seasons. Each prohibition is supported by penalties and enforcement. However, in comparison with the US examples, these institutional innovations by the Kuna have been unsuccessful. Increased pressure on the resource has seen continued decline in yields despite some success in redressing the most egregious acts of over-exploitation. At the time

[6] Ammendments to The Act to Regulate Commerce, The Hepburn Act 1906, The Mann-Elkins Act 1910, The National Transportation Act 1920.

of publication, Sandner argues the continued viability of the resource is in doubt.[7]

What should one make of these observations? They suggest that sometimes new trading opportunities can lead to changes that test the limits of existing institutions. In order to make trade work, new institutions need to be developed. In some situations, this happens and trade is supported, but in others, institutions fail, norms of good behavior are undermined, and trade peters out. After a brief review of relevant literature, the paper's first task is to provide a framework which can explain how both positive and negative outcomes can occur when new trading opportunities arise; Sections 2 and 3. This framework rests on the two key assumptions already stated. The paper's second task is to use the framework to help understand which features of the trading environment, the technology of enforcement, and the parameters of the production technology incline economies to the positive outcome and which to the negative. This is done in Section 4. Finally, the framework is used in Section 5 to help explain why a history of institutional success tends to beget success in new situations, and why institutions that have been developed in successful economies will generally not be usefully transplanted to economies without similar histories.

1.1 Previous literature

As far back as Aristotle, it has ben realized that by repeatedly rewarding good acts, good institutions may actually habituate the opportunistic towards virtue.[8] The institutions of focus here are those that underpin economic transactions which according to Dixit (2006, p.3) comprise three categories: [1] protection of property rights against theft, [2] enforcement of voluntary contracts and [3] provision of the physical and regulatory infrastructure to facilitate economic activity and the functioning of the first two categories. Dixit also distinguishes between relatively formal institutions, such as contract enforcement, the courts and police, and informal self-enforcing institutions: group based punishment that depends on information sharing, repeated interaction, informal sanction or exclusion. Elsewhere, Dixit (2004), emphasizes changes in the prevalence of each at differing stages of development. The transition from the informal to the formal is the focus of Li (2003), Leukaert (2005), Dixit (2004 chs. 2 and 3), Dhillon and Rigolini (2006) and Greif's (2002) analysis of the European community responsibility system. Leukaert (2005) is closest to the present work. He uses this distinction to explain why institutional transplant across countries may fail despite the superiority of the institution being transplanted without the requisite supporting institutions. A type of network externality leading to a scale economy in informal institutional capacity plays a key role there.

The formal/informal distinction central to those analyses plays no significant role here. Here the term "institutions" can be interpreted in both the formal and the informal guises. For example, consider a group of traders attempting to ensure suppliers do not cheat by sharing information about supplier behavior in

[7] See Sandner (2003) for further details.

[8] According to Aristotle's Nicomachean Ethics "Lawgivers make the citizen good by inculcating habits in them, and this is the aim of every lawgiver; if he does not succeed in doing that, his legislation is a failure." See Leukaert (2005, p.2) for this quote and others suggesting a similar process.

a repeated setting. Suppose also that this is relatively ineffective as members of the group are not able to reliably and rapidly communicate. Then, actions taken to improve communication – e.g. employing intermediaries, keeping detailed records, posting and maintaining a common internet site – are all actions that improve institutional quality. The role of a single institution designer may be harder to identify in such settings, but the present model applies provided that key assumption (2) of the introduction continues to hold: some individual, or subset of individuals, is rewarded for improvement efforts, and their rewards are positively related to the value of trade transacted.

Much recent work has looked at the effect of institutions on development, but comparatively little examines the means by which institutions of enforcement are created.[9] Traditional economic approaches to institutional change emphasized the role of competitive forces in driving the evolution of institutions. Demsetz (1967) emphasizes the role of actors searching for the ideal cost minimizing institutions. As Libecap (1986, 227-228) summarizes: "Competitive forces tend to erode institutions that no longer support economic growth. Changing market conditions exert pressure for dynamic adjustments in the existing rights structure through refinements of rights and privileges or their transfer to others...." This is also the essence of the mechanism in Ruttan and Hayami (1984), where political entrepreneurs weigh up marginal costs and benefits to incremental institutional change.

North (1990) extends the standard economic approach, focusing on transaction costs – the costs of policing the trading arrangements – but also emphasizing contingencies that are relevant here. He argues that transactions costs are lower when individuals share beliefs, norms, ideology and a common sense of fairness, which can limit their opportunism. Though noting this, North was not directly concerned with exploring the dynamic connections between norms and institutions; the focus of the present paper.

Greif and Laitin (2004) are also directly concerned with the process of institutional change. Institutions are stable if they are self-enforcing, in the sense of ensuring that underlying parameter values stay close to those that which generated the intitial actions that comprise the institution. Institutions for which this is not true will necessarily change. Though similarly interested in institutional change, the set up here is very different as norms, and changes in them, play a pivotal role.

Legal scholars have also been interested in the role of norms in helping the law – see Cooter (2000) and McAdams and Rasmusen (2006). The latter study reserves the term norm for actions that have a normative requirement, norms are thus always "good". Here, I allow for both good and bad norms to emerge, and assume they are sustained when they confer pecuniary advantages on their followers. To my knowledge, the present paper is the first to formally explore the role of norms, which are endogenous, in affecting endogenous institutional change.

[9] Acemoglu, Johnson, Robinson (2001,2002) are directly concerned with finding causative relationships from institutions to economic outcome variables. Greif (2006) is an exception in that his book is concerned with the process of institution formation. Much of what follows here, particularly in modeling the effects of history in Section 5, departs from insights in his work.

2 The Model

2.1 The Environment

"Entrepreneurs" – There exists an infinitely elastic supply of profit-maximizing, risk neutral producers. They can freely enter into an exogenously given trading opportunity – a technology – or consume their reservation utility, normalized to zero. If entering into the trading opportunity, the entrepreneur's output is sold on a competitive market. With z units of output sold in aggregate, the price of a single unit is given by the inverse demand function $p = AV(z)$, denoting the willingness to pay of the marginal consumer at output level z. A is an exogenously given parameter, with $V(z) > 0$, $V'(z) < 0$ and Inada conditions to ensure an interior equilibrium $V(0) \to \infty, V(z \to \infty) = 0$. Production requires the investment of a fixed, sunk cost C, to produce one unit of output. Production is one shot, and time is continuous. In summary, production yields something of value p (which is endogenous) at cost, C (exogenous and sunk).

2.1.1 One Sided Moral Hazard

"Trading Partners" – Entrepreneurial production requires a trading partner. Working with the trading partner is essential for output to be produced but the trading partner can steal all of the output and sell it on the market at price p. For simplicity, the costs of filling a trading partner's position are set to zero, which is also the opportunity cost of trading partners, and there are no additional costs to cheating. There is no possibility of supporting honest behavior on the part of a payoff maximizing trading partner other than through the formation of institutions, described below.

2.1.2 Formal Institutions

Institutions detect and punish cheating, in particular, a trading partner's stealing of output. Institutional quality at time t is the probability that such theft is detected; $I(t) \epsilon [0, 1]$. There are no false detections.[10] Institutions range from those that are perfect and therefore always detect cheating, $I(t) = 1$, to those that are completely ineffective and never do; $I(t) = 0$. Detected cheaters are punished an amount P, which we treat as a parameter of the institutional technology. The entrepreneur always loses output when cheated. That is, even if the cheater is detected and punished by the formal institution, no portion of the entrepreneur's output will be returned.[11]

[10] Here institutional improvement simply improves the rate of detection when cheating, since there are no errors arising from punishing non-cheaters. A simple extension would include false guilty accusations, in which case institutional improvement would also amount to reducing these. This would not seem to change any of the model's substantive results.

[11] The cost of operating the institution is not modelled here. Introducing this would not greatly impinge on the set-up here, except in requiring some means of taxing and financing the running. It would, however, not affect the comparative statics which are the main focus. Also, it has already been analyzed elsewhere, see Gradstein (2004), and is hence omitted.

2.1.3 First Best Output

Since the only cost borne in production is the cost C, the first best level of entry is given by N solving:

$$\begin{aligned} p &= C \\ \Rightarrow AV(N^*) &= C \\ \Rightarrow N^* &= V^{-1}\left(\frac{C}{A}\right). \end{aligned} \tag{1}$$

2.1.4 Rational agents and norm followers

Proportion $r > 0$ of the set of potential trading partners are rational agents. These agents always act to maximize their payoffs, taking into account expected costs and benefits. They will steal output if the cost benefit ratio is low enough. The remaining proportion, $1 - r > 0$ are norm followers. There are two possible norms. These individuals can either follow a norm of honesty, which dictates that they always trade in good faith and do not steal output, or they can follow a cheating norm and always steal output. At time t, denote the fraction of norm followers following an honesty norm $\beta(t) \in [0, 1]$, with the remaining $1 - \beta(t)$ being cheaters. r is exogenous and β is endogenously determined below. An individual's type – honest, cheater or rational – is known to the individual but not observable to others.

2.2 Dynamics

I first outline the forces leading to change in the quality of institutions. The next sub-section describes the means by which the population's norms change.

2.2.1 Institutional Dynamics

Institutions change via the efforts of a risk neutral institutional designer. This individual is rewarded for institutional improvements, as detailed below, and decides how much effort to devote to improving institutional quality. Such efforts are denoted e and these effort costs are given by the increasing convex function $c(e)$, with $c(0) = 0$.

The designer is rewarded for institutional improvements $\dot{I}$, by the traders on both sides of the market: the producers out of profit, and the consumers out of surplus. The most natural interpretation of the institutional designer is a political actor who is rewarded for improvements in the trading environment evidenced by improved detection, prosecution and punishment of traders who violate agreements. The rewards may take the form of increased political donations, a perception of competence on the part of constituents that increases re-election probability, some other benevolent career concerns, or even direct transfers to the designer. The designer's myopia may be literal, in the case of political agents with short-lived mandates and re-election concerns, or reflect difficulties in appropriability due to incomplete markets.

For instance, as would occur when future beneficiaries of institutional development cannot be identified and taxed to reward the institution building efforts incurred today.[12] This simple way of treating institutional change is a first pass at the problem which clearly abstracts from many important elements of reality. A richer analysis of the political economic considerations that are abstracted from here is left to future work.

The reward function is of a two part form which depends on the current state of institutional performance. If an institution is fully functional, so that it is already effective in deterring rational agents from cheating, then further institutional improvements raise neither welfare nor surplus and the planner receives no reward for these. If institutions are not fully functional, then the designer is rewarded for improvements in institutional quality. The rewards are denoted R and these potentially vary depending on the level of consumer, Ψ, and producer surplus, Π. It has previously been argued that, because producers are easier to organize than consumers, they may be more able to demand regulatory improvements.[13] In order to capture this possibility, it is necessary to allow R to potentially vary with the side of the market that benefits, i.e.:

$$R = F(\dot{I}, \Psi) + G(\dot{I}, \Pi),$$

with $F(0, \Psi) = G(0, \Pi) = 0$ for all Ψ and Π, $F_i \geq 0$, $G_i \geq 0$, for $i = 1, 2$, and $F_{12} \geq 0, G_{12} \geq 0$. That is, instantaneous rewards are higher the greater the institutional improvement, the higher is consumer or producer surplus, for a given institutional improvement, or the larger the degree of improvement at a given level of consumer or producer surplus.

The amount of institutional improvement is directly increasing in the amount of effort contributed by the designer according to the concave technology represented by the function Z:

$$\dot{I} = Z(e),$$

with, $Z(0) = 0$, $Z' > 0, Z'' \leq 0$.

Assuming a simple quasi-linear form for the institutional designer's utility function, the designer's problem becomes:

$$\begin{aligned} & \max_e R - c(e) \\ = \quad & \max_e F(Z(e), \Psi) + G(Z(e), \Pi) - c(e) \end{aligned}$$

This problem leads to a well defined interior solution denoted e^* obtained from the following first order

[12] It is assumed that agents who would actively benefit from continued dysfunctional institutions – for instance those whose aim is to cheat good faith trading partners – have no capacity to influence the institution designer. Hoff and Stiglitz (2004) analyze a case when this can occur to explain the lack of institutional improvement in post-Soviet Russia. Agents whose aim it is to strip corporate assets or to defraud share-holders for personal gain, will oppose (or buy off) institutional designers who aim to improve corporate governance and accountability. Since the costs of maintaining institutional quality are also not modeled here, institutional dynamics are simple and monotonic.

[13] See Posner (1974) for an early statement of the rudiments of this argument. This also underlies formal political economy models that allow for differential treatment of the effects of consumer and producer welfare on political decision makers – as in, for example, Grossman and Helpman (1994).

condition:

$$(F_1(\cdot) + G_1(\cdot)) Z'(e) = c'(e). \tag{2}$$

The solution e^* is weakly increasing in both Ψ and Π, under the assumed conditions $F_{12}, G_{12} \geq 0$. For notational convenience we shall write the solution as $e^*(\Psi, \Pi)$ as defined in (2). Institutional dynamics are thus given by

$$\dot{I} = Z(e^*(\Psi, \Pi)). \tag{3}$$

2.2.2 Behavioral Dynamics

The most simple treatment of behavioral dynamics supposes that, when violating norms pays well, individuals will be tempted to do so, leading to the erosion, and eventual breakdown of the norm in question. In contrast, when following a norm is profitable, the behavior is copied and reinforced. Here I adopt Bisin and Verdier's (2001) model of cultural selection which parsimoniously captures such a treatment. This model has become a widely used way of treating cultural evolution in economics and has received a number of different applications: preferences for social status (Bisin and Verdier, 1998), corruption (Hauk and Saez-Marti, 2002), hold up problems (Olcina and Penarubbia, 2004), development and social capital (Francois and Zabojnik, 2005), intergenerational altruism (Jellal and Wolf, 2002), labor market discrimination (Saez-Marti and Zenou, 2004), globalization and cultural identities (Olivier, Thoenig and Verdier, 2005) and work-ethics (Bisin and Verdier, 2005a), see Bisin and Verdier (2005b) for a full survey of the literature using this approach.[14]

The total population of norm followers is constant at $1 - r$. At time t, each member of the current population socializes a single member of the successive population that will follow it at time $t + \Delta t$. Denote the socializing agent a "parent" and the agent socialized as an "off-spring". There are two ways that socialization occurs in this model. The first, called direct socialization, happens when the parent socializes its offspring to be the same type as itself. The probability of direct socialization is denoted d^i, where $i = h$ or c, for an honest or cheater type respectively. A parent does not have the capacity to directly socialize an off-spring in a way that is counter to type, but with probability $1 - d^i$ the off-spring does not receive direct socialization from the parent. When off-spring are not directly socialized, then socialization is determined by some other individual in the population. The individual of influence is randomly drawn from the remaining population members. This has the appealing feature that, in a society where most people are type i, a person not directly socialized by a parent, will have a high chance of being socialized to be type i by someone else.[15]

[14] The reduced form of behavioral dynamics would be very similar if, instead of a model of cultural evolution, a model where parents make direct choices about childrens' values were to be used. Such a model has been used by Lindbeck and Nyberg (2006) to analyze the effect of welfare provision on work norms.

[15] Rational agents, who recall occur with probability r, are unable to socialize either way. So the probability of indirect socialization depends only on the distribution of other norm followers. An additional layer of selection between rational and non-rational agents could be analyzed provided some costs to rationality were also modeled. This is not attempted here, though the evolutionary stability of pay-off maximizing behavior has been the subject of a substantial literature. For a recent contribution see Heifetz, Shannon and Speigel (2004). A recent contribution by Benabou and Tirole (2005) analyzes the role of incentives in motivating behavior when pro-social considerations are present. That analysis of individual decision making

We are thus able to define the probability that a parent of type i will have an offspring of type j by the function q^{ij}. With proportion $\beta(t)$ honest types at time t, this yields the following four transition probabilities:

$$\begin{aligned} q^{hh} &= d^h + \left(1 - d^h\right)\beta(t) \\ q^{hc} &= \left(1 - d^h\right)(1 - \beta(t)) \\ q^{ch} &= (1 - d^c)\,\beta(t) \\ q^{cc} &= d^c + (1 - d^c)(1 - \beta(t)). \end{aligned}$$

To calculate $\dot{\beta}$ consider $\beta(t)$ and the transition to $\beta(t + \Delta)$ in time interval Δ. Thus we have:

$$\beta(t + \Delta) = \beta(t)\, q^{hh}\Delta + (1 - \beta(t))\, q^{ch}\Delta.$$

Substituting from above, rearranging and letting the time interval $\Delta \to 0$ yields:

$$\dot{\beta}(t) = \beta(t)(1 - \beta(t))\left(d^h - d^c\right) \tag{4}$$

Hence, the two key variables driving the dynamics are the probability that an honest norm follower directly socializes an off-spring to be honest, d^h and the same probability for a cheater, d^c.

The most reasonable assumption to make here is that consistent with a replicator dynamic: the higher relative returns to following an honesty (cheating) norm the higher the probability of direct socialization to honesty (cheating). Since the net benefit of being honest is $PI(t) - p(t)$ we have:

Assumption (Replicator): $d^h - d^c = \phi(PI(t) - p(t))$, with $\phi' > 0$.

2.3 Actions

2.3.1 Trading Partners

Entrepreneurs, at each time t, are randomly matched to trading partners drawn from a population with r rational types and $\beta(t)$ proportion of the non-rational honest, with the remaining proportion $1 - \beta(t)$ cheaters. To recap, honest norm followers never cheat, while cheaters always do. The rational remainder will do so if the expected returns to cheating are higher. Given institutional quality $I(t)$, the price of output $p(t)$, and punishment P, the rational will cheat if and only if

$$p(t) - I(t)\, P > 0. \tag{5}$$

is much more sophisticated than here, as it is focused on the indvidual. Here, instead, the simpler treatment of individuals facilitates our focus on institution formation.

2.3.2 Entrepreneurs

Given norm observance $\beta(t)$, given institutional quality $I(t)$, and given output value, $p(t)$, entrepreneurs will enter into production if and only if the expected returns to doing so are non-negative. That is:

$$p(t) = AV(N(t)) \geq \begin{cases} \frac{C}{r+\beta(t)(1-r)} \text{ when (5) fails} \\ \frac{C}{\beta(t)(1-r)} \text{ when (5) holds} \end{cases}. \tag{6}$$

When (5) holds, only those following an "honesty" norm trade honestly. Given $\beta(t)$ entrepreneurs then enter up to N:

$$\begin{aligned} AV(N(t)) &= \frac{C}{\beta(t)(1-r)} \\ &\Rightarrow \quad N(t) = V^{-1}\left(\frac{C}{A\beta(t)(1-r)}\right). \end{aligned} \tag{7}$$

Let the solution to this be denoted $\widetilde{N}(\beta)$. Under the assumptions on $V(\cdot)$, $\widetilde{N}$ is unique and increasing in β.

When (5) fails, the rational also trade honestly. We can then solve for $N(t)$ as:

$$\begin{aligned} AV(N(t)) &= \frac{C}{r+\beta(t)(1-r)} \\ &\Rightarrow \quad N(t) = V^{-1}\left(\frac{C}{A(r+\beta(t)(1-r))}\right). \end{aligned} \tag{8}$$

Let the solution to (8) be denoted $\widehat{N}(\beta)$, which is again increasing in β and unique.

2.4 Evolution of Norms and Institutions

2.4.1 The Competitive Case

We first analyze equilibria assuming unrestricted entry of entrepreneurs. In such a competitive market, free entry dissipates ex ante profits so that $\Pi = 0$.

Institutions

Institutional evolution depends on the institution designer's effort. Designer efforts depend on the returns to institutional improvement. Specifically from part 2.2.1 we have that :

$$e^* \begin{cases} = 0 \quad \text{when (5) fails} \\ \text{solves } (F_1(\cdot) + G_1(\cdot)) Z'(e) = c'(e) \quad \text{when (5) holds.} \end{cases} \tag{9}$$

When (5) fails, no rational agents cheat, so that institutional improvements never affect behavior and they do not generate rewards for the institution designer. Consequently since $R = 0 \Rightarrow e^* = 0$. When (5) holds, institutions are not yet effective enough to dissuade rational agents from cheating, so that the institution designer is rewarded for institutional improvements according to the functions, $F(\cdot) + G(\cdot)$ as elaborated above.

When the amount of output sold is N, consumer surplus is given by:

$$\Psi = \int_0^N AV(s)\,ds - pN.$$

Since the amount sold on the market is increasing in β irrespective of whether (5) holds or fails, Ψ is necessarily increasing in β too, (the standard area beneath the demand curve which rises with the amount of trade). It is also clear that as $\beta \to 0$, $\Psi \to 0$. For notational simplicity, we shall ignore the other arguments and denote the consumer surplus as simply a function of β by $\Psi(\beta)$. When (5) holds, we can substitute for $\widetilde{N}$, defined in (7), and impose zero profits so that $p = \frac{C}{\beta(t)(1-r)}$, to obtain:

$$\Psi(\beta(t)) = \int_0^{\widetilde{N}(\beta(t))} AV(s)\,ds - \frac{C}{\beta(t)(1-r)}\widetilde{N}(\beta(t)). \tag{10}$$

This is the only case in which consumer surplus matters since, when (5) fails, institutions are already sufficiently good to dissuade the rational from cheating, and further improvement will not increase compliance nor rewards to the designer. Substituting this solution into (9) and recalling that $G(\cdot,0) = 0$, (so that necessarily $G_1(\cdot,0) = 0$) yields

$$\widetilde{e}\begin{cases} = 0 & \text{when (5) fails} \\ \text{solves } F_1(Z(e),\Psi(\beta))\,Z'(e) = c'(e) & \text{when (5) holds.} \end{cases}$$

Substituting into (3) yields the equation of motion for insititutions:

$$\begin{aligned} &\dot{I}\begin{cases} = Z(0) = 0 \text{ when (5) fails} \\ = Z(\widetilde{e}) \quad \text{when (5) holds,} \end{cases} \\ &\textit{where } \widetilde{e} \textit{ solves } F_1(Z(e),\Psi(\beta))\,Z'(e) = c'(e) \\ &\textit{and } \Psi(\beta) \textit{ is given in } (10). \end{aligned} \tag{11}$$

2.4.2 Norms

Given Assumption (Replicator) above, norms evolve according to:

$$\begin{aligned} \dot{\beta}(t) &= \beta(t)(1-\beta(t))\,\phi(PI(t) - p(t)) \\ &= \beta(t)(1-\beta(t))\,\phi(PI(t) - AV(N(t))). \end{aligned}$$

The rate of change in norms thus depends on the expected returns to cheating versus honesty:

$$\dot{\beta}(t) = \beta(t)(1-\beta(t))\begin{cases} \phi\left(PI(t) - AV\left(\widehat{N}(t)\right)\right) & \text{when (5) fails} \\ \phi\left(PI(t) - AV\left(\widetilde{N}(t)\right)\right) & \text{when (5) holds.} \end{cases} \tag{12}$$

3 Equilibrium

There are two notions of equilibrium that vary by time horizon.

(i) Static instantaneous equilibria — holding $I(t)$ and $\beta(t)$ fixed, a mutually consistent levels of entrepreneurial entry, $N(t)$, and decisions for rational trading partners.

(ii) Technology specific steady states — for a given technology the points (I, β) such that $\dot{I} = 0$, and $\dot{\beta} = 0$.

In analyzing such steady states, it is also of interest to track the transition path from a given starting value of honesty norm following β_0, dictated by (11) and (12). In a later section, we will also explore the determination of β_0 as a function of the technologies introduced upto time t.

3.1 Instantaneous Equilibria

For a given technology at time t, the environment is described by an instantaneous level of institutional quality, $I(t)$, and an instantaneous level of norm following, $\beta(t)$. At any such instant, in an instantaneous equilibrium, entrepreneurial entry satisfies constraint (6) and rational agents cheat if and only if (5) holds. The following proposition shows that this yields three possible types of instantaneous equilibria:

.

Proposition 1 *Given $\beta(t), I(t)$, we either have:*
(1) A Poor Institutional Environment. For $I(t) < \frac{C}{(r+\beta(t)(1-r))P}$ the unique pure-strategy equilibrium involves entrepreneurial entry satisfying $AV(N) = \frac{C}{\beta(t)(1-r)}$, and rational partners cheat.
(2) A Good Institutional Environment. For $I(t) \geq \frac{C}{\beta(t)(1-r)P}$ the unique pure strategy equilibrium inolves entrepreneurial entry of N satisfying $AV(N) = \frac{C}{r+\beta(t)(1-r)}$, and rational partners are honest.
(3) An Ambiguous Institutional Environment. For $I(t) \in \left[\frac{C}{(r+\beta(t)(1-r))P}, \frac{C}{\beta(t)(1-r)P}\right)$ there are two pure strategy equilibria. One corresponding to 1. and one corresponding to 2.

When institutional quality is low, rational agents will cheat and entrepreneurs only expect to realize output when matched with honest norm followers; situation (1) above. The converse holds in situation (3). However, there is a strategic complementarity in the actions of rational agents that makes cheating (non-cheating) self reinforcing over certain ranges. Specifically, pairs of beliefs (over equilibrium actions) and actions are mutually reinforcing. Within the range $I(t) \in \left[\frac{C}{(r+\beta(t)(1-r))P}, \frac{C}{\beta(t)(1-r)P}\right)$, when it is believed by entrepreneurs that all rational agents will not cheat, then the high level of entry lowers the marginal product of entrepreneurship and implies that returns are low for any single entrepreneur. In that case, the amount that can be stolen by cheating is low, and rational agents will not cheat. However, when it is believed by entrepreneurs that all rational agents will cheat, then few enter and this makes the marginal value of

output – and hence the return from cheating – high. Consequently, the low level of entry reinforces cheating which reinforces the beliefs underlying the low level of entry. Instantaneous equilibria in the (I, β) space are depicted below.

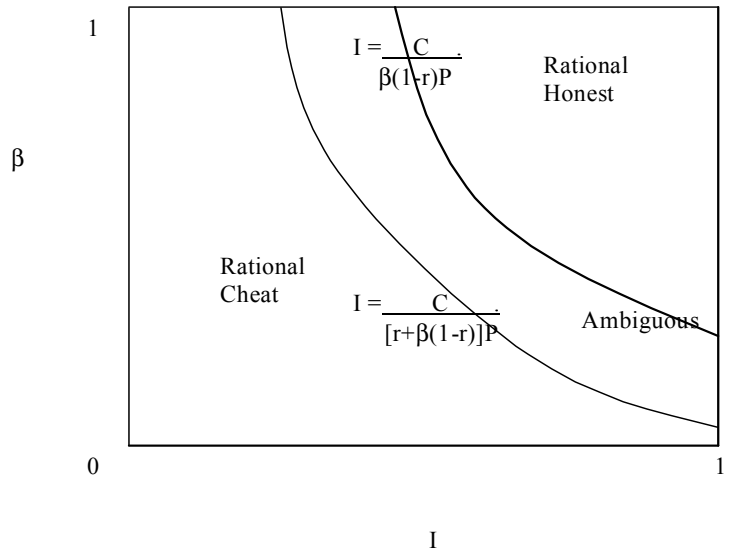

Figure 1. Instantaneous Equilibria

3.2 Play in the Ambiguous Region

In the ambiguous region, pure strategy equilibria either have functioning institutions and no cheating, or the entirely opposite dysfunctional institutions and cheating. It turns out that none of the qualitative results to follow in this paper depend on how this multiplicity is resolved, i.e., on whether it is assumed that the cheating or the honest equilibrium is played. It is therefore without loss of generality that we assume that, when in the region with multiple equilibria, the cheating equilibrium is played. As will be seen below, given the dynamics of this system, this will seem like a more natural focal equilibrium, but to reiterate, none of the qualitative results depend upon it.[16]

3.3 Technology Specific Transition Paths and Steady States

3.3.1 Steady States

Formal institutions never get worse, $\dot{I} \geq 0$. Internalized norms get worse whenever the rational have incentive to cheat, and improve whenever the rational choose to be honest. This makes the set of possible steady state outcomes relatively small:[17]

[16] A detailed demonstration of this is available from the author upon request.
[17] We ignore the set of unstable steady states, in the usual tatonnement sense.

Proposition 2 *There are two types of stable steady state in the model:*

1. Fully functional institutions: Steady states where there is no cheating and output is at the first best level N^*. *These lie in the range:* $I \geq \frac{C}{(1-r)P}$ *and* $\beta = 1$.

2. Dysfunctional institutions: Steady states where production shuts down entirely, and cheating would be certain if production were to be attempted. These lie in the range $I \in [0, 1]$ *and* $\beta = 0$.

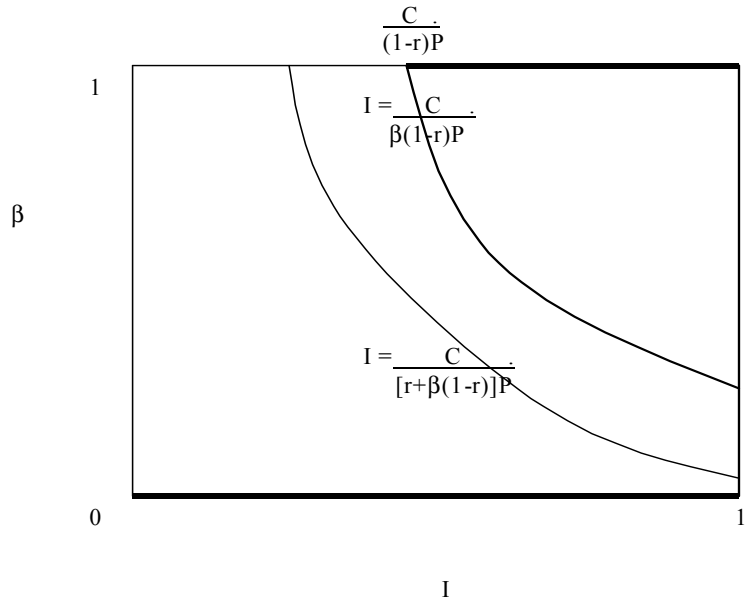

Figure 2. Steady States (Bold)

In situation 1 of the proposition, all norm followers are honest, and cultural evolution also favors honest norms. Trade is at the first best level so there are no further rewards for institutional improvement and they remain as is. In situation 2, all non-rational agents are cheaters. Were there to be entry by an entrepreneur, rational trading partners would always cheat, since the expected returns to cheating are large and outweigh the expected costs of punishment – even when punishment is certain. Consequently, there is no entry into production and trade breaks down completely. As the returns to cheating dominate those to honesty, cultural evolution continues to favor cheating.

3.4 Transition Paths

The phase space of the system is given by (11) and (4) and is extremely simple. When institutions are not fully functional, below the line $I = \frac{C}{\beta(1-r)P}$, institutions are not good enough to dissuade the rational from cheating, and cheaters do better than honest types. The honesty norm is eroded and cheating increases, β falls. Institutions are improving, however, as the institution designer is rewarded for institutional improvement in this region. The system evolves in a South/Easterly direction in this region. Above the line

$I = \frac{C}{\beta(1-r)P}$, returns to honesty exceed those to cheating, as institutions are good enough to dissuade the rational from cheating. In that case, institutions are fully functional, and no further institutional improvement is required. Honesty is reinforced by good institutions and β increases.

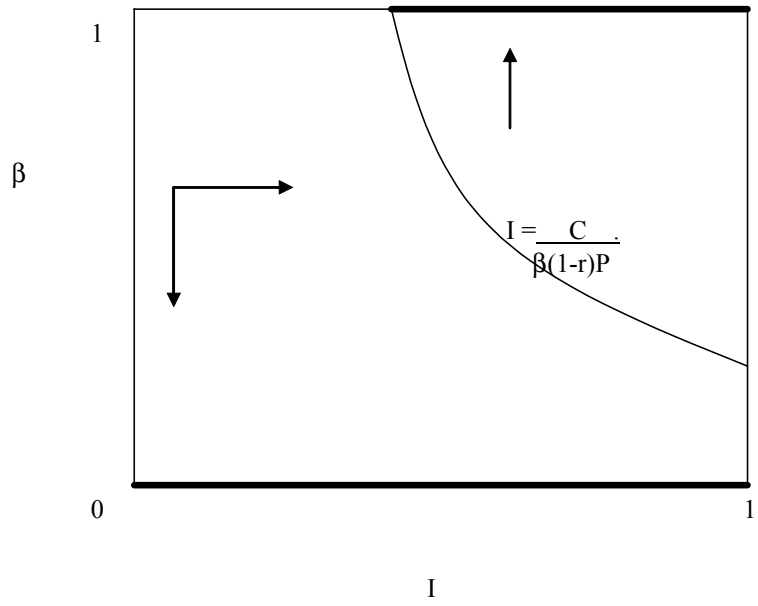

Figure 3. Phase Space

3.4.1 Critical Starting Values

Clearly, the higher is the initial level of honesty norm following, β_0, ceteris parabus, the more likely is the system to converge on a steady state with functional institutions. Define a critical level of initial honestly norm following that is just sufficient to ensure the development of functional institutions.

Definition: *Given initial institutional quality $I = 0$, denote β_0^c as that value of β such that if and only if $\beta_0 \geq \beta_0^c$ then the economy develops functioning institutions and converges to a steady state where no cheating occurs.*

The appendix demonstrates that such a critical level is unique and well defined when it exists, and when it does not exist there is no possibility of an economy transitioning to functional institutions. The figures below show the critical level under different scenarios. The corresponding value of I at the point where institutions become functional is denoted I^c. In what follows, we will be interested in characterizing how β_0^c is affected by changes in the parameters of the environment.

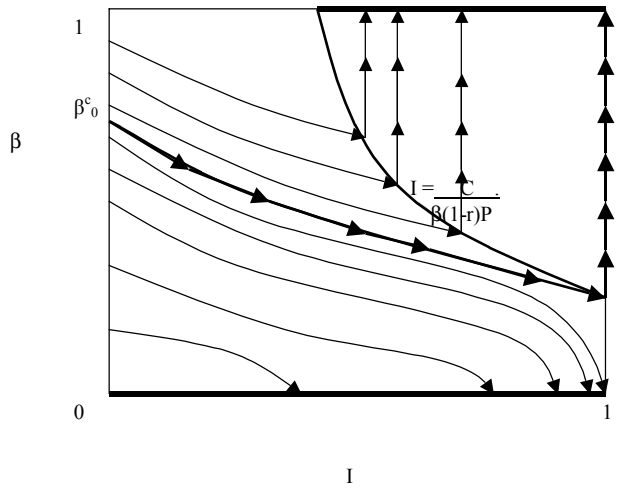

Figure 4A. Transition Paths ($I^c = 1$)

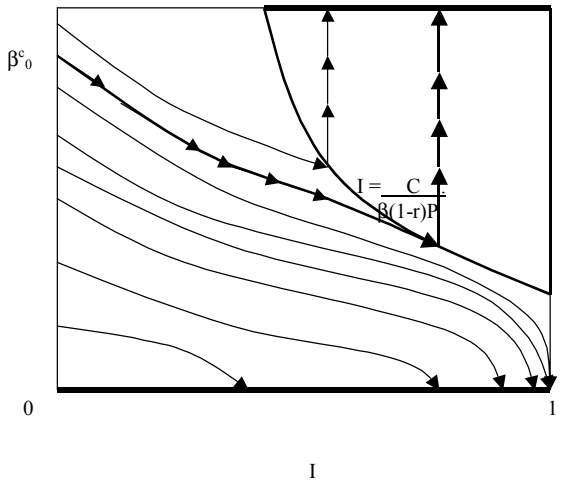

Figure 4B. Critical ($I^c < 1$)

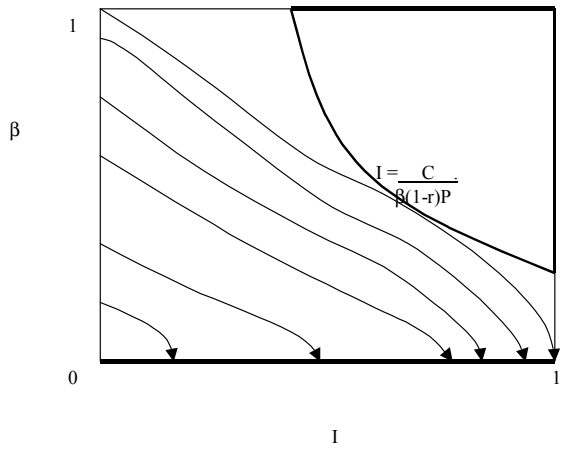

Figure 4C. No β_0^c defined

4 Comparative Dynamics

The following proposition characterizes the effect of variations in model parameters on β_0^c and on the required level of institutional quality sufficient to ensure functional institutions.

Proposition 3 *(1) (Productive Technologies) Increasing A does not affect the level of institutional quality required to ensure incentive compatibility but lowers the starting level of honesty norm compliance needed to ensure functional institutions – β_0^c falls).*
(2) (Costly Technologies) Increasing C raises the level of institutional quality required to ensure incentive compatibility, and raises the starting level of honesty norm compliance needed to ensure functioning institutions – β_0^c rises.

(3) (Punishment) Increasing P lowers the level of institutional quality required for functioning institutions and lowers the starting level of honesty norm compliance needed to ensure functioning institutions – β_0^c falls.

A higher value for A implies that the slope of the expected transition path, $\frac{d\beta}{dI}$ is less negative. Thus it is more likely that successful institutions will be attained in the case of more productive technologies. Intuitively, before institutions are fully functional, a more productive technology implies that, at any given β, greater numbers of entrepreneurs enter production. The higher N implies greater surplus from trade, greater reward for institutional improvement, and hence more effort from the designer to do so. Faster institutional development implies that a lower level of initial honest norm compliance will be sufficient for functioning institutions to be attained.

Increasing C has the opposite effect. In order to defray higher costs of entry, marginal productivity must be higher in instantaneous equilibrium, implying less entrepreneurial entry. But with less entrepreneurial entry, returns to cheating are higher so that a correspondingly higher level of institutional quality is needed to ensure rational agents will not cheat – the $I = \frac{C}{\beta(1-r)P}$ shifts up, thus shrinking the region where institutions function. Along the expected transition path, the lower value of N also implies that returns to institutional improvement are everywhere lower. Moreover, since returns to cheating are higher everywhere, good norm erosion is faster. Consequently the slope of the expected transition path, $\frac{d\beta}{dI}$, is steeper (more negative). The initial level of good norm following required to ensure successful institutions will be attained is therefore higher.

With greater punishment there is less inducement to cheating, thus a lower level of institutional quality needs to be reached to ensure compliance. Also, there is less good norm erosion because net expected returns for cheaters along the transition path are everywhere lower. For a given β, the level of entrepreneurial entry is the same, as is the amount of cheating, so the rate of learning and institutional adaptation is unchanged. In net, the trajectory is flatter implying a lower critical level of norm following to reach the same level of institutional quality.

4.1 Restricted Entry

Up until now we have assumed that entrepreneurial entry is unrestricted. Entrepreneurs enter until the point where expected profits are zero. We now explore the effects of entrepreneurship being limited by vested interests. This is clearly the case in many developing countries – for example, India's license Raj, or system of permits and regulations required for entry into industrial sectors prior to the reforms of the early 90s.[18] We ask what effect, if any, such a restriction has on the level of institutional quality required to ensure compliant trading partners, and on the likely development paths to such institutional quality.

[18] The World Bank's "doing business" web-site, www.doingbusiness.org, provides detailed estimates of the magnitude of these restrictions for many developing countries. This is also consistent with Parente and Prescott (2002) emphasis on vested interests acting to thwart competitive entry.

Assume now then that entrepreneurial entry is restricted to a subset of potential entrepreneurs of size $\overline{N}$. The only interesting case is where this is binding for at least some situations. Since entry is maximal when institutions are sufficiently good to ensure the rational comply, a necessary condition for this to bind is that:

$$AV\left(\overline{N}\right) > C. \tag{13}$$

We assume this condition holds for $\overline{N}$ in what follows. An implication is that, in any part of the space where the entry restriction binds, profits Π are positive.

Let $\overline{\beta}$ denote the value of β:

$$\overline{\beta} = \frac{1}{(1-r)}\left(\frac{C}{AV\left(\overline{N}\right)} - r\right),$$

from (13) $\overline{\beta} < 1$. This is the value of β above which, given $\overline{N}$, profits are positive. Now, the planner may also be rewarded from profits as well as consumer surplus. Where $\overline{N}$ does not bind, producer surplus, $\Pi = 0$, as before, and consumer surplus is $\Psi(\beta)$, as defined in equation (10). Where $\overline{N}$ binds, then both consumer and producer surplus are functions of $\overline{N}$ as well as β.

When the entry restriction binds, and where institutions are not functional, the price of output equals $p = AV\left(\overline{N}\right)$, and total profit $\Pi = \overline{N}\beta(1-r)p - C\overline{N} = \overline{N}\left(\beta(1-r)AV\left(\overline{N}\right) - C\right)$. Consumer surplus is: $\Psi = \int_0^{\overline{N}} AV(s)\,ds - \overline{N}AV\left(\overline{N}\right)$, which is straightforward to show is increasing in $\overline{N}$. In contrast to the previous, now consumer surplus is only a function of $\overline{N}$, not β, since entry is bound and unaffected by β, we thus denote it: $\Psi\left(\overline{N}\right)$. It is easy to verify that Π is increasing in β, while its response to $\overline{N}$ depends on the elasticity of demand. We denote it by $\Pi\left(\overline{N},\beta\right)$ in this restricted entry case.[19]

Where institutions are not functional, e is chosen to satisfy:

$$\left(F_1\left(Z(e), \Psi\left(\overline{N}\right)\right) + G_1\left(Z(e), \Pi\left(\overline{N},\beta\right)\right)\right) Z'(e) = c'(e).$$

If $F(\cdot,x) = G(\cdot,x) = F(\cdot,y) + G(\cdot,z)$ where y and z are any pairs satisfying $y + z = x$ then the distribution of producer and consumer surplus do not matter for the rewards that the planner receives. We will call this the case where consumers and producers have "equivalent influence" on institution designers. In that case, all that matters for rewards to the producer is the total surplus, $\Pi + \Psi$, and not its distribution. We shall contrast this with another case where producers are more influential than consumers, that is where: for all $x > 0$, $F\left(\dot{I}, x\right) > G\left(\dot{I}, x\right)$ for any $\dot{I} > 0$.

Proposition 4 *(License Raj) Restricting entrepreneurial entry to $\overline{N}$ reduces the (β, I) region where institutions are functional. If producers have equivalent influence on institutional designers as consumers, then restricting entrepreneurial entry also requires a higher level of initial norm compliance for any given level of institutions to be attained – β_0^c rises.*

[19] The sign is $\beta(1-r)A\left(V\left(\overline{N}\right) + \overline{N}V'\right) - C$, which is ambiguous.

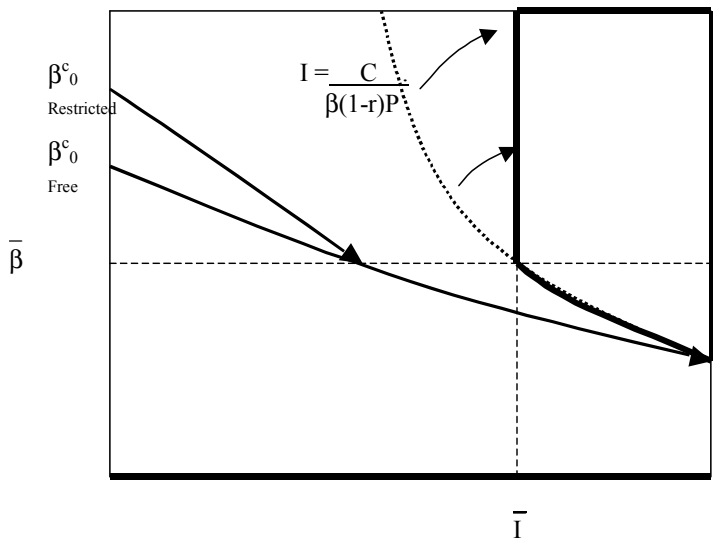

Figure 5. Restricted Entry

By restricting entry, the $\overline{N}$ firms obtain rents from production. Consequently where institutional quality ensured compliance without entry restriction, for some points, it will no longer be assured. The higher rents leave a greater amount to be stolen and require higher quality institutions. This rotates the upper part of the region required for functional institutions above $\overline{\beta}$ until it lies on the line $I = \overline{I}$. Additionally, along the transition path, whenever the entry restriction binds (that is above $\overline{\beta}$) the trade surplus is lower. Under the "equivalent influence" hypothesis, where all that matters is the level of trade surplus, and not its distribution, this has the dual effect of both lowering the amount of effort put into institutional improvement, and increasing the returns to cheating, so that honesty norm erosion is faster. The transition path is then steeper in this region so that higher initial levels of honesty norm following are needed to develop any given level of institutional quality.[20] Both forces thus make functional institutions less likely to develop with restricted entry.

Increases in Productivity

Recall from Proposition 3 part 1, that when entry was unrestricted, a more productive technology was more likely to lead to the development of functional institutions. In some ways, that result resembled one set of concerns emphasized by North (1990) – the institutions enabling the more productive technologies to work generate the highest social payoffs, and are thus naturally more likely to receive the efforts of institutional designers. We now show that this is not true when entry is restricted:

[20] If and only if $\beta^c_0 < \overline{\beta}$, then there is no change in β^c_0 under restricted entry. In contrast, the region where institutions are functional always shrinks under restricted entry.

Proposition 5 *(License Raj 2) When entrepreneurial entry is restricted, to* $\overline{N}$*, a more productive technology (increase in* A*) reduces the* (β, I) *region where institutions are functional. If producers have, at most, "equivalent influence" as consumers on institutional designers, the level of initial norm compliance needed to attain functional institutions is ambiguous.*

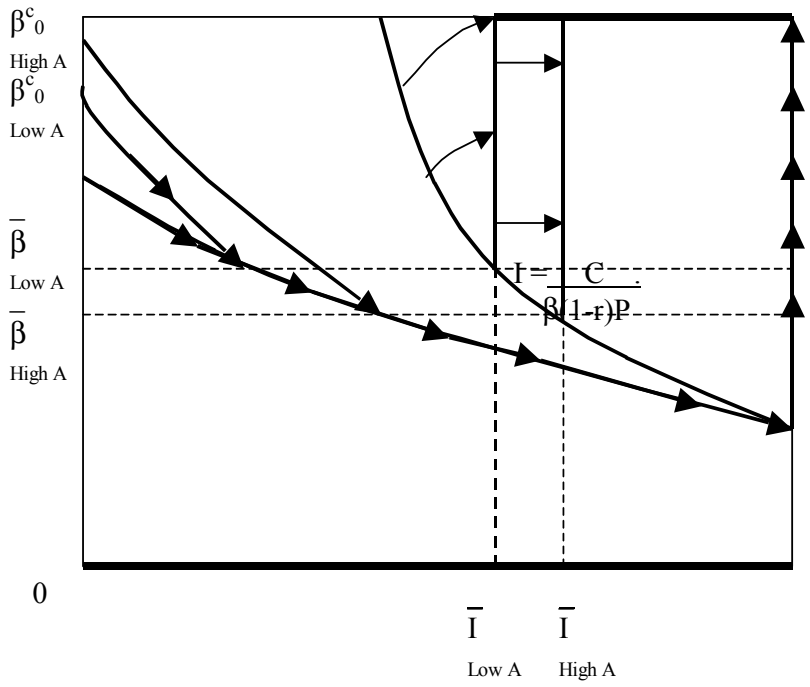

Figure 6. Restricted Entry (High A)

Some levels of institutional quality that ensured compliance with low A will now not do so under higher A. This is because, when $\overline{N}$ binds, more productive technologies ensure yet more can be stolen by cheating trading partners. Consequently, the already reduced region in which institutions function when entry is restricted, is further restricted when a technology is more productive – the functional region shrinks from the left – as depicted by the horizontal arrows in the figure above.

The effect on transition path is less clear. On the one hand, the entry condition binds wherever it did under the less productive technology, but now at a higher marginal product – ensuring the returns to cheating and norm erosion are now greater in the binding region. On the other hand, the increase in A raises returns to institutional improvement, ensuring faster improvement ceteris parabus. The former effect makes the path flatter while the latter makes it steeper so that, overall, the change is ambiguous.

Note the difference from the unrestricted entry case. In that case, higher productivity was dissipated

by entry so that the level of institutional quality required was unaffected. Also, the higher entry ensured greater surplus to trade and ensured that institutional improvement along the transition path was everywhere greater.

Producer Influence Exceeds Consumer Influence

Now suppose that, since producers are more easily organized, surplus in their hands will have greater impact on institution design than in the hands of consumers. Under this treatment, the earlier result, arguing that restricted entry makes the development of functional institutions less likely, no longer holds. This now becomes ambiguous because the restrictions also transfer surplus from consumers to producers and, if producers are influential enough (i.e., $G >> F$), then the institution designer may be more greatly rewarded for improvements, even though the overall level of surplus generated by trade is lower. Under these conditions, restricting entry could help in the formation of functional institutions by benefiting the set of constituents who are more effective in rewarding the institution designer. But, though the dynamics of transition towards functional institutions could conceivably be improved with restricted entry, it remains the case that the region of the phase space in which institutions are functional, always shrinks when trade is restricted. Moreover, the shrinkage remains greatest when A is high, i.e., when technologies are most productive. This implies that even in this most optimistic case for restricted entry we have the following:

Proposition 6 *Even if restricting entry helps in improving the dynamic paths towards the creation of functional institutions, for highly productive technologies, ($A > \frac{P}{V(\overline{N})}$), under free entry, functional institutions will be developed, but with restricted entry dysfunctional institutions are the only outcome.*

Restrictions on entry can only help in forming functional institutions for relatively low productivity technologies. Even under the most optimistic conditions – when skewing the distribution of surplus in favor of producers increases rewards to institutional designers, and hence the speed of adjustment – it remains the case that, for high productivity technologies, restrictions on entry unambiguously impede the formation of functional institutions. Intuitively, highly productive technologies and restricted entry create high producer rents so that returns to cheating can never be adequately mitigated by improved enforcement. With such huge returns to malfeasance, even when these create a force favoring rapid institutional improvement, institutions can never be good enough to ensure that rational individuals will have incentive to trade honestly. Under free entry, in contrast, such returns would have been dissipated to consumers, making the surplus unable to be stolen by opportunistic traders and hence allowing the formation of functional institutions.[21]

[21] Note that this result is mitigated to an extent if the designer is able to increase P, the punishment to caught cheaters. However, the proposition still holds for any environment in which P is finite.

5 Steady States for Multiple Technologies

In this section we explore the impact of an economy's past experience with earlier technologies on its likelihood of reaching functioning institutions with a new one, and on the level of institutional quality required to do so. In order to do this, we must introduce multiple technologies. Suppose now that, at time t, an economy has been subject to the introduction of $M(t)$ new technologies, each of which requires institutions supporting trade to be developed. Now the subscript i denotes the ith technology. Assume that these are in horizontally differentiated sectors, and in each one, functional institutions must be designed anew. Thus at time t, any available technology is denoted by an $i \in \{1, M(t)\}$. Let the steady state levels of institutions and behavioral norms reached for each one of these technologies be given by I_i, β_i.

For simplicity we shall assume that technologies are introduced sequentially, with the subsequent technology introduced only after the previous one has reached its steady state. Let $S(t)$ denote the number of technologies having reached a steady state with $\beta_i = 1$ and consequently functional institutions upto time t. The remaining $M(t) - S(t)$ have reached steady states where $\beta_i = 0$, and institutions are not functional.

Societies are intertemporally linked by the starting levels for norm following behavior which are a function of past experience. As Greif (2006) has argued, agents "search for coordination drawing on the past, and seek consistency of behavior with internalized norms and beliefs....institutional elements inherited from the past are part of the initial conditions of processes that lead to new institutions." p. 194

What do these "institutional elements" from the past do? "In uncoordinated processes leading to new institutions, past institutionalized norms and beliefs provide individuals with a means of forming expectations about others' behavior and hence deciding on their own behavior. When a new situation is similar (by whatever criteria) to an old situation, individuals expect that others will behave in the new situation as they did in the old one. Self enforcing beliefs formed in one context constitute a focal point (Schelling 1960) that coordinated behavior in a new one."p 203.

In short, according to Greif whose lead we shall follow from hereon, when new situations arise, where existing norms of behavior are not yet well established, individuals tend to act in ways that correspond with the norms of behavior that comprise the steady states of existing institutions. These are denoted by the existing steady state level of β_i for those trading situations that have already been experienced. Hence we assume:

Assumption *Conformism – When behavioral norms are not yet established in a new trading situation, agents base their actions on the observed behavior of others in existing trading situations. That is, the proportion of agents initially following an honesty norm for a technology introduced at time t, denoted* $\beta_0(t)$ *is increasing in the ratio* $\frac{S(t)}{M(t)}$.

When most previous trading opportunities reach a steady state where all individuals comply with trade

honestly – $\frac{S(t)}{M(t)}$ is high – then, in new situations, individuals will also tend to start off with honesty norms; $\beta_0(t)$ will be high. In contrast, where many technologies reach steady states where trade fails, and it is expected that new traders would be cheated almost certainly, then, in new situations, relatively few individuals will follow an honesty norm. We assume the possibility of a stochastic component to this increasing relationship as follows:

$$\begin{aligned}
\beta_0(t) &= g\left(\frac{S(t)}{M(t)}\right) + \mu \\
g' &\geq 0 \\
g &: 1 \to 1 \\
\mu &\sim \ i.i.d, \\
E[\mu] &= 0 \\
\sigma^2[\mu] &> 0.
\end{aligned}$$

We shall, for now, assume that there is no possibility of institutional learning from another source, say from the experiences of another country, so that all technologies continue to commence at institutional level $I = 0$. This shall be relaxed subsequently to explore the possibility of institutional transfer across countries.

The assumption above does not pin down the level of norm following that a society starts with at the advent of its very first trading opportunity, denoted $\beta_0(0)$, but it does imply considerable path dependence from then on. Specifically:

Proposition 7 *Consider any two economies A and B such that the level of norm following for the first trading opportunity in economy A exceeds that for the first one in economy B; $\beta_0^A(0) > \beta_0^B(0)$. Then, economy A is more likely to attain successful institutions in technology 1 and subsequent technologies than economy B.*

If the value of $\beta_0(0)$ is high, an economy is more likely to achieve successful institutions for the first trading opportunity than is an economy starting with a low value of $\beta_0(0)$. But then, at time t when the second technology becomes available, since $S(t) = 1$ for a successful economy, and 0 for an unsuccessful one, $\beta_0(t)$ for the successful economy is likely to lie above that for the unsuccessful. It will once again be more likely for the successful economy to converge on functional institutions. But this will make it even more likely still that the economy will develop functional institutions subsequently, and so on. Of course, a set of poor realizations of the stochastic variable can upset this path, but there is, at least in expectation, strong path dependency.

5.1 Institutional development and the legacy of history

It is clear then that, under the assumed avenue of positive feed-through into initial conditions, a past marked by considerable institutional failure directly affects the likelihood of developing successful institutions in future. However, a less immediate implication is that the level of institutional quality required to generate functional institutions is higher for a country with a history of failed institutions. Specifically:

Proposition 8 *Consider two economies A and B where A has a more successful institutional history than B at a given time. The level of institutional quality required for functional institutions in country B is higher than the quality required for economy A.*

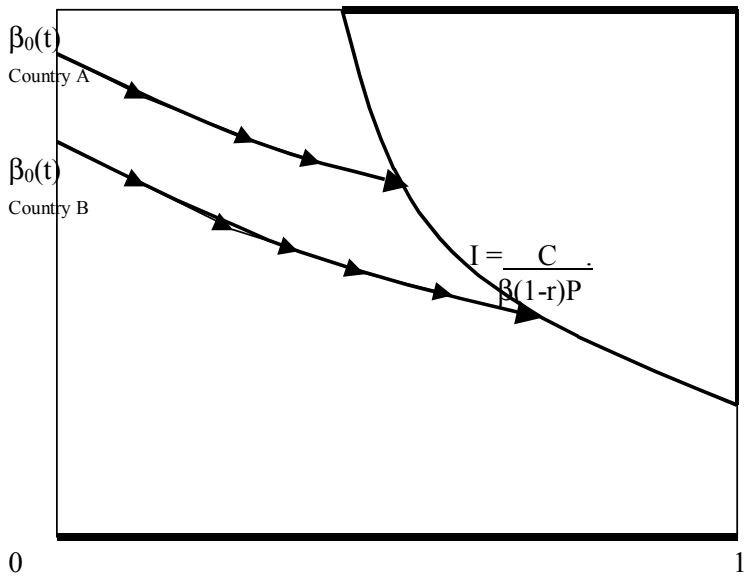

Figure 7. Functioning Institutions Higher in B

The fact that an economy starts with a lower level of the good norm implies that, in the event that it does attain successfully functioning institutions, it will do so for a lower final value of β, and higher I. Since it must support functioning institutions at a lower β, the level of entrepreneurial entry at any point in time will be lower in the country with a history of failed institutions. Consequently, incentives to steal will be higher (as the price of output is higher), so that institutions will have to be higher quality to ensure that the rational will not do so. Thus, conditional upon generating a functional institution in a particular case, we should expect that an economy with a poor history of institution formation will have a higher quality institution than one with a successful history.

We now explore the effects of transferring institutional knowledge from a country with a history of institutional success, and hence discovery of functioning institutions, to another with a history of failed institution formation.

Proposition 9 *(Institutional Transplant) Suppose that country B with a history of institutional failure is able to perfectly and costlessly learn from the steady state quality level of institutions from successful country A at time t. (1) This will increase the likelihood that country B develops a successful institution for a new trading technology at t. (2) The greater the divergence in historical experience, i.e. in the ratios* $\frac{S^i(t)}{M^i(t)}$ *the more likely that the imported level of institutional quality that functions in country A will NOT be sufficient to ensure functional institutions in country B.*

Countries with similar histories will be more able to share each others institutional successes. Specifically, if Country B has a similar institutional experience to A, any success in A will almost certainly lead to functioning institutions in B.

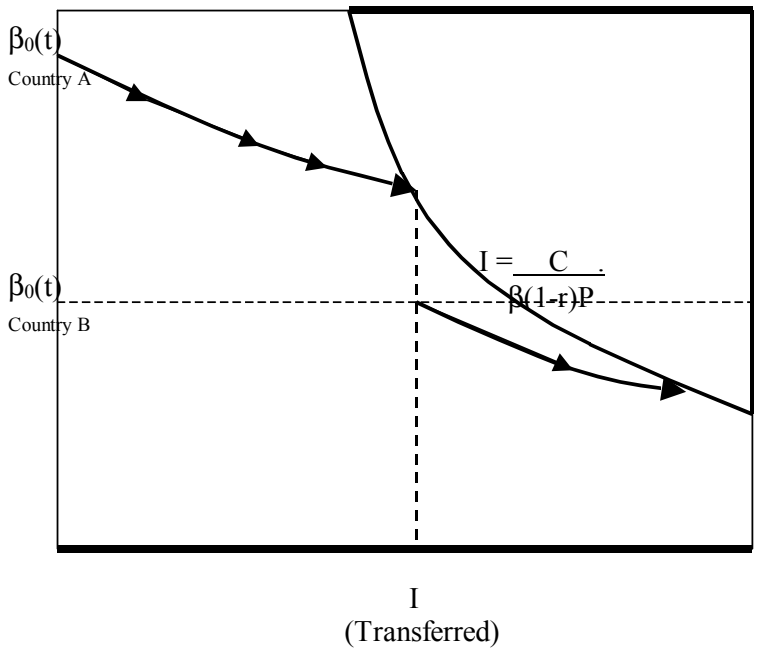

Figure 8. Transfer of Institutions

But the institutional past of successful countries will be valuable for counties without such histories too. To the extent it is transferrable, it gives them a leg-up in the design of their own institutions for the relevant trading situation. However, even if the institutional knowledge is fully transferrable it will not generally be enough to ensure institutional success in the recipient economy. This is because the lower level of norm following in the historically unsuccessful one implies it is possible to sustain a bad equilibrium in which

rational agents cheat for higher values of I. This possibility is increased the greater the divergence in history – implying the larger difference in β_0 between countries. In order to preclude this possibility in the less successful economy, more institutional knowledge than that created in the successful economy is required. This can only be acquired in the recipient country since institutional designers in the successful country have no incentive to develop their own institutions beyond the level necessary for them to function at home. Countries with a history of dysfunctional institutions will learn more from countries with similar histories, conditional upon these others creating successful institutions, than they would from countries that have been successful.

6 Conclusions

Individuals can be good, but relying on them to be good will not work for long if incentives do not also line up. There is a sense in which good institutions make good agents – as has long been suggested – but good agents are also necessary for good institutions to be able to arise in the first place.

This paper has taken many short cuts in its analysis of the interaction between norms and institutions. It models a very stylized political process of institution development, it assumes that the running of functioning institutions is costless, it assumes that norms also change in a stylized way, and that agents who would gain from creating poor institutions have no role in the political process. It also assumes that agents' types are binary, and producers are homogeneous. But all of the results here will survive some degree of tampering with these assumptions. The key assumptions upon which results critically depend are: (1) Good institutions create good (norm following) people, bad ones do the opposite, and both of these processes take time; (2) Good institutions take time to build, and the incentive to do so depends on the current gains from the trade that the institution is designed to regulate.

The main finding here is that restrictions to entry may severely limit the development of functional institutions. The first reason for this is that it reduces the gains from trade. Since it is this surplus to trade which is ultimately drawn upon to reward institutional designers, incentives for institutional improvement are lower, ceteris parabus, when entry is restricted. But this ceteris parabus caveat is particularly limiting in the present context, as restricting entry also has distributional consequences which can offset the lowering of surplus effect. Specifically, restricting entry shifts surplus from consumers to producers. If producers have greater influence on institutional designers than consumers, then even though the total surplus is lower, since more accrues to producers, institutional designers can have greater incentive to develop functional institutions.

There is, however, a more subtle damaging effect of restricting entry. Restricting entry raises the value of the marginal unit produced. This also means that it raises the value of the marginal unit which can be stolen. Consequently it raises the level of institutional quality necessary to ensure that opportunistic

traders will trade honestly. Restricting entry thus increases the demands made on institutional design, and these demands are greatest the more productive is the technology. Even if it is the case that skewing the distribution of returns towards producers helps in providing incentives for institutional designers, it is ALWAYS the case that for sufficiently productive technologies, economies with entry restrictions will not be able to attain functional institutions whereas if they had allowed free entry they would have done so.

The final section of the paper considered the possibility of transferring institutional knowledge between countries. Specifically, countries that had not yet succeeded in finding functioning institutions for a particular technology were allowed to freely and perfectly import the institutions already developed to do the same task in a successful country. Moreover it was assumed that the environment, and hence institutional requirements, in both countries were identical. Somewhat surprisingly, this free and perfect transfer of identical institutions was not generally found to work. It was, in fact, most likely to fail when the transfer was from a country with a history of creating functional institutions to one which had a history of institutional failure. Though transferred knowledge was always useful, it was not likely to be sufficient, and the greater the difference in the history of institutional success in recipient and source country, the more deficient was the knowledge likely to be. Countries with a history of failed institutions are more likely to gain by importing the subset of successful institutions (if any)of countries with a similar history to theirs, than by importing them from countries with successful pasts.

An important assumption underlying this analysis is that the individuals capable of changing institutions have a direct interest in their improvement. This is because they are directly rewarded out of the surplus generated by the trade that institutions support. At a first pass, this seems like the right assumption to make. But economic theory is also rich with models of more complex political economic interactions. Reality also seems to be littered with examples of institution designers who have thwarted the development of institutions because their capacity to extract surplus has been tied to existing, inefficient, institutional forms. The present framework would also s seem suited to exploring such cases, which is the task of future work.

References

Acemoglu, D., S. Johnson and J. Robinson (2001) The Colonial Origins of Comparative Development: An Empirical Investigation, American Economic Review, 92, 1369-1401.

Acemoglu, D., S. Johnson and J. Robinson (2002) Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution, Quarterly Journal of Economics, 118, 1231-1294.

Bartley, T. (2003) Certifying Forests and Factories: States, Social Movements and the Rise of Private Regulation in the Apparel and Forest Product Fields, Politics and Society, 31, 3, 2003, 433-464.

Benabou, R. and J. Tirole (2005) Incentives and Prosocial Behavior IZA, Bonn, discussion Paper #1695.

Bisin, A. and T. Versider (1998) On the Cultural Transmission of Preferences for Social Status, Journal

of Public Economics, 70 (1), 75-97.

Bisin, A. and T. Verdier (2001): 'The Economics of Cultural Transmission and the Dynamics of Preferences,' Journal of Economic Theory, 97, 298-319.5.

Bisin A. and T. Verdier (2005a), 'Work Ethic and Redistribution: A Cultural Transmission Model of the Welfare State,' mimeo, NYU

Bisin, A. and T. Verdier (2005b) Cultural Transmission, New Palgrave Dictionary of Economics, Second Edition.

Chandar, N., and P. Miranti (2005) Institutions, information and agency: the crisis of railroad finance in the 1890s and the evolution of corporate oversight responsibilities, Accounting History.

Coleman J., (1988), 'Social Capital in the Creation of Human Capital,' AmericanJ ournal of Sociology, 94, S95-S120.

Cooter, R.D. (2000) Do good laws make good citizens, Virginia Law Review, 86(8) 1577-1601.

Demsetz, H. (1967) Towards a theory of property rights, American Economic Review.

Dhillon, A. and J. Rigolini (2006) Development and the Interaction of Enforcement Institutions, mimeo, Warwick University, UK.

Dixit, A. (2004) Lawlessness and Economics: Alternative Modes of Governance, Princeton University Press, Princeton NJ.

Dixit, A. (2006) Economic Governance, New Palgrave Dictionary of Economics, forthcoming.

Fafchamps, M. (2004) Markets and Institutions in Sub-Saharan Africa: Theory and Evidence, Cambridge, MA, MIT Press.

Francois P. and J. Zabojnik, (2005), Trust Social Capital and the Process of Economic Development, Journal of the European Economics Association.

Gradstein, M. (2004) Governance and Growth, Journal of Development Economics, 73, 505-518.

Greif, A. (2002) Institutions and Impersonal Exchange: From Communal to Individual Responsibility, Journal of Institutional and Theoretical Economics, 158(1) 168-204.

Greif, A. (2006) Institutions and the Path to the Modern Economy: Lessons from Medieval Trade, Cambridge, Cambridge University Press.

Greif, A. and D. Laitin (2004) A theory of endogenous institutional change, American Political Science Review, 98, 4, 1-20.

Grossman, G and E. Helpman (1994) Protection for Sale, American Economic Review, 84(4) 833-50.

Hauk E. and M. Saez-Martı (2002), 'On the Cultural Transmission of Corruption,'Journal of Economic Theory 107(2), 311-335.

Heifetz,A. C. Shannon, Y. Speigel (2004) What to maximize if you must, UC Berkeley Working Paper.

Hoff, K. and J. Stiglitz (2004) After the Big-Bang? Obstacles to the Emergence of the Rule of Law in

Post-Communist Societies, American Economic Review, 94(3) 753-63.

Jellal, M. and F.C. Wolff (2002), 'Cultural Evolutionary Altruism: Theory and Evidence,' European Journal of Political Economy, 18, 241-62.

Leukeart, A. (2005) The Dynamics of Institutional Change: Formal and Informal Institutions and Economic Performance, mimeo, Munich.

Li, J.S. (2003) Relation Based versus Rule Based Governance: An explanation of the East Asian Miracle and Asian Crisis, Review of International Economics 11, 651-73.

Libecap, G. (1986) Property Rights in Economic History: Implications for Research, Explorations in Economic History, 23, 227-252.

Lindbeck, A and S. Nyberg (2006) Raising Children to Work Hard: Altruism, Work Norms and Social Insurance, Quarterly Journal of Economics, forthcoming.

McAdams, R.H. and E. Rasmusen (2006) Norms in Law and Economics, forthcoming Handbook of law and economics Vol. 2, A. Polinsky and S. Shavell editors, Amsterdam, Elsevier.

North, D. (1990) Institutions, Institutional Change and Economic Performance, Cambridge, Cambridge University Press.

Olcina G. and C. Penarrubia (2004), 'Hold-Up and Intergenerational Transmission of Preferences,' Journal of Economic Behavior and Organization, 54, 1, 111-132

Olivier J., M. Thoenig and T.Verdier (2005), 'Globalization and The Dynamics of Cultural Identity,' mimeo, PSE

Parente, S. and E. Prescott (2002) Barriers to Riches, MIT Press, Cambridge.

Patacchini, E. and Y. Zenou (2005), 'Intergenerational Education Transmission:Neighborhood Quality and/or Parents' Involment?,' mimeo, IUI

Posner, R.A. (1974) Theories of Economic Regulation, Bell Journal of Economics and Management Science, 5(2), 335-358.

Ruttan, V.W. and Y. Hayami (1984) Towards a theory of Induced Institutional Innovation, Journal of Development Studies, 20, 203-223.

Saez Marti' M. and Y. Zenou (2004): 'Cultural Transmission and Discrimination,'mimeo, IUI

Sandner, V. (2003): Myths and Laws: Changing Institutions of Indigenous Marine Resource Management in Central America, in How Institutions Change edited by H. Breit, A. Engels, T. Moss and M. Toja published by Leske and Burdich, Opladen, 2003

Schelling T.C (1960) The Strategy of Conflict, Harvard University Press, Cambridge MA.

Vranceanu, R. (2003) Manager Unethical Behaviour During the New Economy Bubble, ESSEC Cergy Pontoise, working paper 03026.

Appendix

Proof of Proposition 1: Part 1. When $I(t) < \frac{C}{P(r+\beta(t)(1-r))}$, then even at $N' : AV(N') = \frac{C}{r+\beta(t)(1-r)}$, $AV(N') - PI(t) > 0$. Consequently for any $N \leq N'$, since V is a decreasing function $AV(N) - PI(t) > 0$ and it is worthwhile for rational agents to cheat. Entry of $N > N'$ is never possible, so rational agents will always cheat in this case. Since rational agents cheat, entrepreneurial entry must satisfy: $(1-r)\beta AV(N) = C$.

Part 2. When $I(t) > \frac{C}{\beta(t)(1-r)P}$ then for any value of $\beta(t) < 1$, and $\widetilde{N}$ defined in the text such that $AV\left(\widetilde{N}\right) = \frac{C}{\beta(t)(1-r)}$, it follows directly that $AV\left(\widetilde{N}\right) - PI(t) < 0$, so that rational agents would not cheat under $\widetilde{N}$. However, if rational agents do not cheat, then entrepreneurial entry, N', must satisfy $AV(N') = \frac{C}{r+\beta(1-r)}$. And since $V(N') < V\left(\widetilde{N}\right)$ it is also the case that $AV(N') - PI(t) < 0$. Thus rational agents will not cheat under N'.

Part 3. When $I(t) \in \left[\frac{C}{(r+\beta(t)(1-r))P}, \frac{C}{\beta(t)(1-r)P}\right)$ we demonstrate that there exist two mutually reinforcing sets of beliefs regarding actions, and hence two pure strategy equilibria. We ignore mixed strategy equilibria. Equilibrium a) N' and no cheating. If it is believed that all rational agents will not cheat, entrepreneurs enter upto N' where $AV(N') = \frac{C}{r+\beta(t)(1-r)}$. With this level of entry, rational agents will not cheat if and only if

$$\frac{C}{r+\beta(t)(1-r)} - PI(t) \leq 0$$

which holds in this range. Consequently entry of N' and not cheating is a self-reinforcing pair.
Equilibrium b) If it is believed that all rational agents will find it worthwhile to cheat when I is in this range, then entrepreneurs enter upto $AV\left(\widetilde{N}\right) = \frac{C}{\beta(t)(1-r)}$. It will be worthwhile for agents to indeed cheat provided that:

$$\begin{aligned} AV\left(\widetilde{N}\right) - PI(t) &> 0 \\ & \quad\; \textit{equivalently} \\ \frac{C}{\beta(t)(1-r)} - PI(t) &> 0. \end{aligned}$$

Which holds in this range. Consequently rational agents cheating is consistent with the entry of entrepreneurs in the presence of cheating. ■

Proof of Proposition 2: Values of $\beta \in (0,1)$ can only satisfy (12) for $I = \frac{C}{\beta(1-r)P}$. But at these value of I, small perturbations will move the system away from these points, (eigenvalues > 0). If $\beta = 0$ then clearly $I < \frac{C}{\beta(1-r)P}$, and $\dot{\beta} < 0$, but it is bounded at 0 and since there is no trade, $e^* = 0$ so that $\dot{I} = 0$. If $\beta = 1$ and $I \geq \frac{C}{(1-r)P}$, then $\dot{\beta} > 0$, but it is bounded at 1 and since institutions are functional, $R = 0 \Rightarrow e = 0 \Rightarrow \dot{I} = 0$.

Solving for β_0^c:

This critical level β_0^c can be solved for in the following three steps: 1) find the lowest path that transitions from $I = 0$ to the locus $I = \frac{C}{\beta(1-r)P}$. If no such path exists, then β_0^c is not defined and there is no possibility

for this economy to transition to functional institutions, this is the case depicted in Panel C of the text. 2) If such a path exists, then at the point where the path meets the line $I = \frac{C}{\beta(1-r)P}$ denote the value of I as I^c. In Panel A of the text, $I^c = 1$, and in panel B I^c <1. 3) Given I^c, β_0^c is the value of β solving the following initial value problem:

$$\begin{aligned} \dot{\beta} &= \beta(1-\beta)\left(PI - AV\left(\widetilde{N}\right)\right) \\ \dot{I} &= F(\widetilde{e}) \\ I(T) &= I^c, \beta(T) = \frac{C}{I^c(1-r)P} \\ I(0) &= 0. \end{aligned} \tag{14}$$

Graphically, this is given by the β_0^c starting at the bold path highlighted with arrows in panels A and B.

Proof of Proposition 3

Part 1. The level of institutional quality required to ensure incentive compatibility is described by the curve $I = \frac{C}{\beta(1-r)P}$, which is clearly unaffected by A.

Consider the transition range where $\beta > 0$ and (5) holds. At any (β, I) value in this range, the transition path is given by: $\dot{\beta} = \beta(1-\beta)\phi\left(PI - AV\left(\widetilde{N}\right)\right)$, and $\dot{I} = Z\left(e^*\left(\int_0^{\widetilde{N}} AV(s)\,ds - \frac{C}{\beta(1-r)}\widetilde{N}\right)\right)$ where e^* solves (9), given the consumer suplus, $\Psi = \int_0^{\widetilde{N}} AV(s)\,ds - \frac{C}{\beta(1-r)}\widetilde{N}$, and $\widetilde{N}$ solves

$$AV\left(\widetilde{N}\right) = \frac{C}{\beta(t)(1-r)}. \tag{15}$$

Compare two technologies A and $A' > A$. Note that for A' (15) must also hold. Since $V' < 0$, $\widetilde{N}$ solving (15) under A is less than $\widetilde{N}'$ solving (15) under A', i.e. $\widetilde{N} < \widetilde{N}'$. Clearly, (15) also implies:

$$AV\left(\widetilde{N}\right) = \frac{C}{\beta(1-r)} = A'V\left(\widetilde{N}'\right). \tag{16}$$

Compare the slope of the transition path for case A', denoted $\left(\frac{d\beta}{dI}\right)'$ with that for A, denoted $\frac{d\beta}{dI}$:

$$\begin{aligned} \left(\frac{d\beta}{dI}\right)' &\equiv \frac{d\beta}{dt}.\frac{dt}{dI} = \frac{\beta(1-\beta)\phi\left(PI - A'V\left(\widetilde{N}'\right)\right)}{Z\left(e^*\left(\int_0^{\widetilde{N}'} AV(s)\,ds - \frac{C}{\beta(1-r)}\widetilde{N}'\right)\right)} \\ &= \frac{\beta(1-\beta)\phi\left(PI - AV\left(\widetilde{N}\right)\right)}{Z\left(e^*\left(\int_0^{\widetilde{N}'} AV(s)\,ds - \frac{C}{\beta(1-r)}\widetilde{N}'\right)\right)} \\ &> \frac{\beta(1-\beta)\phi\left(PI - AV\left(\widetilde{N}\right)\right)}{Z\left(e^*\left(\int_0^{\widetilde{N}} AV(s)\,ds - \frac{C}{\beta(1-r)}\widetilde{N}\right)\right)} = \frac{d\beta}{dI}, \end{aligned}$$

where the second line follows from (16) and the inequality in the third line follows from $\widetilde{N} < \widetilde{N}'$, $\phi' < 0$ and $PI - AV(N) < 0$ at all points in this range. Since $\left(\frac{d\beta}{dI}\right)' > \frac{d\beta}{dI}$ the solution to the initial value problem

for A' is lower than for A. That is β solving the initial value problem (14) for $A', \widetilde{N}'$ is strictly less than β solving the same initial value problem for $A, \widetilde{N}$.

Proof of Part 2. Functioning institutions require $I = \frac{C}{\beta(1-r)P}$ clearly this is increasing in C, implying that a higher level of institutional quality is required to ensure incentive compatibility for any level of β. Consider the transition range where $\beta > 0$ and (5) holds, so that $PI - AV\left(\widetilde{N}\right) < 0$. From (15), for a given β, increasing C implies $\widetilde{N}$ falls. A lower $\widetilde{N}$ implies $V\left(\widetilde{N}\right)$ rises so that $\left(PI - AV\left(\widetilde{N}\right)\right)$ falls, and also the monotonicity of $\Psi = \int_0^{\widetilde{N}} AV(s)\,ds - C\widetilde{N}$ in $\widetilde{N}$ implies $\Psi\left(\widetilde{N}\right)$ falls. Thus the slope of the transition path which is given by

$$\frac{d\beta}{dI} = \frac{\beta(1-\beta)\left(PI - AV\left(\widetilde{N}\right)\right)}{Z\left(e^*\left(\Psi\left(\widetilde{N}\right)\right)\right)},$$

falls (i.e. steeper negative slope). Since the transition path is steeper, the critical initial level, β_0^c, is necessarily higher.

Proof of part 3. The proof of this part proceeds analogously to the part 2 since P enters only through the expression $C/(1-r)P$, consequently the comparative static for P is the converse of that for C.■

Proof of Proposition 4: For $\beta < \overline{\beta}$ (13) does not bind, $\Pi = 0$, and the phase diagram is identical to that already analyzed in the unrestricted entry case. For $\beta \geq \overline{\beta}$, $\overline{N}$ binds so that $\Pi\left(\overline{N}, \beta\right) > 0$. Denote $I : PI = AV\left(\overline{N}\right)$ by $\overline{I}$. and consider the regions where $I < \overline{I}$ so that $PI - AV\left(\overline{N}\right) < 0$ and (5) holds. Let $e^*\left(\Psi\left(\overline{N}, \beta\right) + \Pi\left(\overline{N}, \beta\right)\right)$ denote the solution to $\left(F_1\left(Z(e), \Psi\left(\overline{N}, \beta\right)\right) + G_1\left(Z(e), \Pi\left(\overline{N}, \beta\right)\right)\right) Z'(e) = c'(e)$. In this part of the phase space, the slope of the transition path, denoted $\frac{\overline{d\beta}}{dI}$, is given by

$$\frac{\overline{d\beta}}{dI} = \frac{\beta(1-\beta)\phi\left(PI - AV\left(\overline{N}\right)\right)}{Z\left(e^*\left(\Psi\left(\overline{N}\right) + \Pi\left(\overline{N}, \beta\right)\right)\right)}. \tag{17}$$

Compare this with the slope of the transition path in the unrestricted entry case:

$$\frac{d\beta}{dI} = \frac{\beta(1-\beta)\phi\left(PI - AV\left(\widetilde{N}\right)\right)}{Z\left(e^*\left(\Psi\left(\widetilde{N}(\beta)\right)\right)\right)}. \tag{18}$$

Since $\overline{N} < \widetilde{N}$, clearly $V\left(\overline{N}\right) > V\left(\widetilde{N}\right)$, also if producers have, at most, equivalent influence on institutional designers as consumers, then this implies that $F\left(\Psi\left(\widetilde{N}(\beta)\right)\right) > F\left(\Psi\left(\overline{N}\right)\right) + G\left(\Pi\left(\overline{N}, \beta\right)\right)$ so that $e^*\left(\Psi\left(\widetilde{N}(\beta)\right)\right) > e^*\left(\Psi\left(\overline{N}\right) + \Pi\left(\overline{N}, \beta\right)\right)$. It then follows that $\frac{\overline{d\beta}}{dI} < \frac{d\beta}{dI} < 0$, and the transition path is steeper in this range. Consequently if $\beta_0^c > \overline{\beta}$ then β_0^c is higher under restricted entry. If $\beta_0^c \leq \overline{\beta}$ it is unaffected.

Consider an $I < \overline{I}$, and $\beta > \overline{\beta}$. Since $\beta > \overline{\beta}$, $AV\left(\overline{N}\right) > \frac{C}{r+(1-r)\beta}$, so that entrepreneurs would like to enter but since entry is restricted to $\overline{N}$, for rational types to not cheat necessarily $I \geq \overline{I}$, consequently for functioning institutions at any $\beta \geq \overline{\beta}$ necessarily $I \geq \overline{I}$, so that the region of the phase diagram where $\dot{\beta} > 0$ rotates rightward to the vertical line as depicted in Figure 5.

■

Proof of Proposition 5: From the previous proposition, the left boundary of the region with functioning institutions is given by $\overline{I} = \frac{AV\left(\overline{N}\right)}{P}$. For $A' > A$, $\overline{I}$ increases to the point denoted $\overline{I}' = \frac{A'V\left(\overline{N}\right)}{P}$. $\overline{\beta}$ as defined in the text also falls. Thus the left boundary of the region shifts rightwards reducing the region of the phase space where institutions are functional. For $I > \overline{I}'$ entry restrictions do not bind so that the region for functioning institutions continues to be defined by the function $I \geq \frac{C}{\beta(1-r)P}$.
Now consider a transition path in the region of the phase space where $\overline{N}$ binds. On any such path

$$\frac{d\beta}{dI} = \frac{\beta\left(1-\beta\right)\phi\left(PI - AV\left(\overline{N}\right)\right)}{Z\left(e^{*}\left(\int_{0}^{\overline{N}} AV\left(s\right)ds - C\overline{N}\left(\beta\right)\right)\right)}.$$

But clearly since $\overline{N}$ binds, for $A' > A$ the numerator of this expression falls (i.e. more negatively sloped). But, for a higher A, the total surplus rises and the level of institutional designer effort, which is increasing in the total surplus given by $\left(e^{*}\left(\int_{0}^{\overline{N}} AV\left(s\right)ds - C\overline{N}\left(\beta\right)\right)\right)$, rises. Since this is the denominator of the expression above, the net effect on $\frac{d\beta}{dI}$ is ambiguous. ■

Proof of Proposition 6: From the proposition above, it has already been demonstrated that an increase in A has an ambiguous effect on the slope of any transition paths for $\beta > \overline{\beta}$, and for $\beta \leq \overline{\beta}$, the dynamics are unaffected. The region under which institutions are functional is the area above the function $I = \frac{C}{\beta(1-r)P}$. When entry is restricted to $\overline{N}$, this is further restricted to $I \geq \frac{AV\left(\overline{N}\right)}{P}$. Now consider an $A > \frac{P}{V\left(\overline{N}\right)}$. Under free entry, at such an A the region for functional institutions remains at $I = \frac{C}{\beta(1-r)P}$. But with restricted entry, there does not exist an $I \leq 1$ which can satisfy $I \geq \frac{AV\left(\overline{N}\right)}{P}$. Thus, irrespective of the slope of the transition path, this economy cannot converge to a steady state with functional institutions and trade. The only steady state outcome is a complete failure of trade and dysfunctional institutions.■

Proof of Proposition 8: Follows directly from the fact that country B starts with $\beta_0\left(t\right)$, lower than that for country A, and the function $I = \frac{C}{\beta(1-r)P}$ is negatively sloped. ■

Proof of Proposition 9: All directly inferrable from Figure 8.■