Explaining the Rise of Institutions: Toward a Kirznerian Theory of Repeated Games

Peter Nencka*1

1. Introduction

In the last 20 years, economists following pioneers such as Douglass North and Oliver Williamson have begun to pay attention to the critical role institutions play in economic growth. Daron Acemoglu, who won the John Bates Clark Medal in 2005, in part for his empirical work on institutions, asserts the “main determinants of cross-country variations in per-capita income are differences in economic institutions” (Acemoglu et al. 2004: 1). But what exactly are institutions? Douglass North’s (1990: 2) definition is a good starting point: Institutions “are the rules of the game in a society or, more formally, are the humanly devised constraints that shape human interaction.” They provide the structure agents need to engage in successful social interactions. Institutions come in many types—both the formal, codified structure of law and the informal, self-enforced structure of cultural and moral norms. As Sautet (2008) notes, when formal institutions are in conflict with the underlying informal norms and culture of a people, the formal rules will be too costly to enforce. Further, bad institutions lead to a perversion of the otherwise socially beneficial entrepreneur. There is a robust and well-developed literature on the stifling role bad institutions have on market

* Peter Nencka is a 2011 graduate from the Department of Economics at Beloit College.

1 Thanks to Josh Hall and Eli Blee-Goldman for valuable comments on earlier drafts of this paper and Emily Chamlee-Wright for direction and feedback throughout the entire process. All remaining errors are my own.
entrepreneurs; see, for example, Kirzner (1979) and Baumol (1990). Comparatively little has been written on the role entrepreneurial activity itself plays in the formation of institutions, particularly informal institutions. My purpose here is to present one such approach as an extension of Ken Binmore’s discussion of the evolution of social contracts, infusing his theory with an Austrian focus on entrepreneurship.

In Section 2 I describe Binmore’s (2007) game theoretic account of the rise of institutions. While pointing to the theoretic clarity of the stability and efficiency conditions in his model, I also discuss shortcomings of his choice of an equilibrium selection device. Section 3 introduces the idea of an Austrian ideological entrepreneur drawing on the sketch offered in Storr (2009). I argue that Storr’s conception of an ideological entrepreneur plays a critical role in explaining institutional change, but that his theory can be buttressed by drawing on the literature in evolutionary game theory as represented by Binmore (2007), Ostrom (1990), and others. Section 4 attempts to bridge the gap between the game theoretic and Austrian concepts of institutional evolution by incorporating a version of Storr’s ideological entrepreneur into Binmore’s model. I argue that the resulting model does not contradict the spirit of either of its components and provides a better explanatory fit than both. Section 5 concludes, outlines next steps, and offers a response to one possible methodological objection.

2. Binmore’s Natural Justice and Institutional Change

At first glance, it might be thought that an account of institutional change reduces to a claim about political change. While the study of government institutions clearly holds important insights for the student of development economics, the formation and stability of those government and legal institutions themselves are a function of underlying social norms. A robust set of formal institutions, such as those that define liberal democracy, cannot survive unless the underlying social norms and customs are supportive. Binmore calls these basic, self-enforcing norms a society’s social contract. In this he builds off of the philosophic tradition

---

2 For instance, liberal democracy cannot merely be exported to other cultures without the risk of full-scale rejection. See Coyne (2007) for an economic analysis of the pitfalls of this type of nation-building.
that grounds the legitimacy of an institution in its ability to secure agreement from all its participants in both real and imagined scenarios.

Laws and government practices are relevant to this discussion only inasmuch as they are actually followed. The force of the state does not overwhelm social contracts; it is in fact justified by the contract. As Binmore (2007: 3) notes, “Popes, presidents, kings, judges, or the police are not exempt from the social contract of the society in which they officiate. Far from enforcing the social contract, they derive what power they have from a social convention which says that ordinary citizens should accept their direction.” Binmore’s theory in Natural Justice is an attempt to characterize the evolution of these basic social norms. In his argument, a social contract must satisfy three requirements: stability, efficiency, and fairness. The first two follow from a straightforward game theoretic account, the final from a questionable appeal to a naturalized version of John Rawls’ theory of justice (Rawls, 1971). Before turning to a critique of Binmore’s theory of justice, I summarize the game theoretic tools needed to understand the bulk of his position.

Game theory is built around a simple idea: interdependence. Economic agents often act strategically, varying their actions in accordance with their rational deliberations on the possible actions of other agents. Modeling this interdependence is the purview of game theory. In traditional economic theory, the focus is different—firms and individuals maximize profits and utility relative to given price and income constraints. In models of perfect competition, firms don’t have to worry about what other firms are doing, since all firms are price-takers. But as F.A Hayek (1948) and others have argued, there appears to be very little competition in perfect competition. Game theory is one attempt to restore a study of the market process as we perceive it in real life: full of bargains, negotiations, and strategy. For example, consider the following prisoner’s dilemma, the most well-known concept in game theory:

Two people are arrested on suspicion of plotting to rob a famous art gallery. The detectives investigating the case need a confession from at least one of the criminals, so they offer the thieves a deal: If one confesses and the other does not, the confessor receives one year in jail and the other thief receives five years. If they both confess, they both serve three years. But if neither confesses, the authorities have only circumstantial evidence, and the thieves can be released after two years.
Suppose each player is interrogated separately and simultaneously. We can formalize our story in a payoff matrix of possible strategies (with Player 1’s payoffs listed first) as:

<table>
<thead>
<tr>
<th></th>
<th>P2 Confess</th>
<th>P2 Silent</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 Confess</td>
<td>3 years, 3 years</td>
<td>1 year, 5 years</td>
</tr>
<tr>
<td>P1 Silent</td>
<td>5 years, 1 year</td>
<td>2 years, 2 years</td>
</tr>
</tbody>
</table>

Clearly, if the prisoners were allowed to talk and set up an enforceable contract, they would agree that both would be silent—the lower right cell is the socially optimal position (from the perspective of the prisoners). But supposing their decisions are made simultaneously and without contact with one another, each now has the incentive to confess. Supposing that P1 thinks that P2 will remain silent, P1 has an incentive to confess, reducing his sentence from two years to one year, thereby moving from the lower right cell to the upper right cell. Supposing that P1 thinks that P2 will confess, P1 still has the incentive to confess in order to avoid the five-year sentence he would receive if he were to remain silent. No matter what P1 expects P2 will do, it is always in P1’s interest to confess. P2 faces exactly the same incentives, and thus, will also always confess.

The argument above shows that Confess, Confess is a Nash Equilibrium, or the dominant outcome, of the prisoner’s dilemma, which from the players’ perspective, is sub-optimal. When players confess, they are making the best possible response to their opponent’s possible strategies and have no incentive to unilaterally change their behavior, and thus, social cooperation toward the prisoners’ optimal outcome is impossible.

The dynamics of our prisoner’s dilemma can be positive if our goal is to incarcerate criminals, but they can represent a vexing problem in a context in which social cooperation is needed. We can imagine everyday scenarios in which similar dilemmas crop up. Friends forming a study group might find that while it would be in the interest of the group if everyone worked diligently; each individual has an incentive to shirk and free-ride on the effort of others. So it may turn out that
everyone shirks. Or consider fishermen deciding if they should continue to overfish a limited resource. If they all agree to refrain from overfishing, they will all be better off, as the stock of fish increases. Yet each fisherman has an individual incentive to overfish in the short term, hoping the others will refrain. But in pursuing their individual interest, the fishermen systematically overfish and all are worse off in the long run.

Originally, collective action problems such as these raised serious doubts about the prospect of widespread human cooperation. So-called “tragedy of the commons” situations appeared to be everywhere! Yet both empirical and theoretical results have challenged this original assumption. Ostrom (1990) provides compelling case studies of non-government solutions to collective action problems that lead to sustained cooperation. Behavioral economists, such as Fehr and Schmidt (1999) recommend overhauling traditional decision theory to include “other-regarding preferences” to help explain cooperation observed in lab and field studies. But such a move is potentially problematic, in that it isn’t so much a solution as it is an ad hoc stopgap. Further, such a theoretical move may not be needed. While remaining within a traditional rational choice framework, game theory nonetheless can shed light on why cooperation can persist over time, and even while accepting the fact that cooperation in the one-shot game violates the traditional notion of rationality.

2.1 The folk theorem for repeated games

Recall the story of our prisoners but now relax the assumption that the game is played just once. In fact, assume they need to decide whether to confess or stay silent an indefinite number of times. (If it helps the intuition, suppose they are degenerate criminals and immediately after getting released from jail they try the same crime again and get caught over and over.)

The one-shot prisoner’s dilemma is thus turned into a repeated game. In a repeated game we can speak about meta-strategies, that is, rules that players use

---

3 Hardin (1968) popularized the term “tragedy of the commons” and set off a lively debate on the possibility we would never be able to overcome these coordination problems.

4 Most formulations of the folk theorem require the games be played infinitely. The theorem does not work if the players know when the last game is because they would defect in that game and then reason backwards, defecting in all other games. While infinitely repeated games do not exist in the real world, if people have no knowledge of when the game might end they will tend to act as if the game is infinite—hence the indefinite assumption.
to guide their actions from game to game. Suppose both players adopt what is known as the grim trigger meta-strategy: If you stay silent in game \( n \), I will stay silent in game \( n+1 \), but if you confess in game \( n \), I will confess in every game \( m \), where \( m > n \). In other words, if you cooperate, I will cooperate back for at least the next game. But if you defect and confess (or shirk in the group project or overfish), I will punish you by confessing for every game from then on.

With a meta-strategy in play, a repeated game can result in cooperation, i.e., a Nash Equilibrium of Silent, Silent.\(^5\) Suppose that in Round 1 both players remain silent. To show this is a Nash Equilibrium, we ask whether either of the players has an incentive to unilaterally defect and confess in the next round. If the players continue to remain silent, each would receive \( 2+2+2+2+2+2+2+2+2 \ldots \) years in jail. But suppose P1 decides to deviate and confess in the third round. Then as punishment, P2 would confess starting in the fourth round. In turn, P1 would punish player P2 for the fourth round confession in every subsequent round. So P1’s jail time would now be \( 2+2+1 \) (the round in which he confessed) + \( 3 + 3 + 3 + 3 + 3 \ldots \), clearly less preferable than the original cooperative strategy.\(^6\) Thus neither player has an incentive to deviate from cooperating with the other. Unlike the one-shot prisoner’s dilemma, cooperation is a viable strategy in the repeated version of the game without abandoning rational decision theory. In general, note that the worst off P1 can possibly be (assuming he is playing rationally) is modeled by his payoff above—his best response to P2’s worst punishment. Why could P1 not be any worst off? By asserting he is rational, we have assumed he will always make the best response to P2’s strategy. Knowing this, P2 will select an optimal punishment, in this case defecting, that minimizes the value of P1’s best response. Game theorists call this the minimax condition. The folk theorem shows that any outcome above minimax condition for both players is a stable Nash equilibrium as long as agents care about their future stream of payoffs. If they are at a payoff higher than their minimax condition, they have no incentive to unilaterally defect because they know they could be worse off when punished.

\(^5\) In general, there might be a wide range of meta-strategies, which can lead to cooperation in indefinitely repeated games, such as a strategy in which one only punishes an opponent for one turn following defection. For a general proof that characterizes the conditions meta-strategies must follow: see Ratliff (1996).

\(^6\) Mathematically, both sums approach infinity and thus, in some sense, are equal. But human decision makers have finite lives and thus the differences in payoff streams will be meaningful. Some game theorists avoid the mathematical problem by focusing on average payoffs in repeated games or by redefining utility functions to eliminate the infinite sum.
The folk theorem shows that repeated interaction can foster cooperation and repeated interaction is exactly what characterizes life in a cosmopolitan society. The folk theorem provides a way to judge whether a particular public policy or institution is feasible: does the benefit it provides to those cooperating outweigh the utility those agents could get from defecting and playing their best response to the punishments such defecting entails? If so, we have a self-enforcing institution that requires no government or threat of external force to maintain. For Binmore, the folk theorem provides the set of feasible institutions that could be selected without social cooperation falling apart; hence he calls it his “stability criterion.” His “efficiency criteria” follows quickly from similar reasoning: social contracts must not only be stable within a community, they need to be competitive with other communities. To illustrate, consider one of Binmore’s favorite examples: the driving game (Binmore, 2007: 60). Drivers in a community need to choose whether they will drive on the left or the right. There are three Nash equilibria in this repeated game: 1) everyone drives on the right, 2) everyone drives on the left, 3) everyone flips a coin every time they drive—if it is heads they drive on the right, tails they drive on the left. To see that this last game is a Nash equilibrium, note that if one player is flipping a coin, the other player gains nothing from changing his strategy from coin flipping since he will still get hit 50% of the time no matter what he does. Hence if we start at a coin-flipping state, neither player has an incentive to switch strategies, the definition of Nash equilibria (Binmore, ibid.).

But even though there are three Nash equilibria in this repeated driving game, we do not want to claim that all three are on equal evolutionary footing. The coin-flipping strategy is stable inside a community, in the sense that neither player acting alone has an incentive to defect. Yet compared to a community in which people do not crash 50% of the time, the coin-flipping social contract is clearly unattractive. The relative inefficiency of the coin-flipping social contract predicts it will not last: either people will move to a better community or simply die off. Driving on the right or on the left are both stable, efficient equilibria in Binmore’s terminology. Coin-flipping is stable, yet not efficient. So far, we have seen that a selected social contract must be both a Nash Equilibrium (as characterized by the folk theorem) and competitive. Yet as the driving game shows, these two conditions do not predict what social contract is ultimately selected—there can be potentially limitless stable and competitive equilibria.

When dealing with moral matters, Binmore argues that we choose between these efficient and competitive equilibria by deploying an evolved sense of fair-
ness that roughly corresponds to a naturalized version of John Rawls’ theory of justice. In short, he argues, we have developed a heightened sense of empathy—one that allows us to make interpersonal comparisons of utility by “imagining ourselves in the shoes of others” (Binmore 1998: 56). It is this sense of fairness that eventually leads to the selection of one efficient equilibrium over another. While Binmore’s first two conditions for a plausible social contract—stability and efficiency—merely show what forms of cooperation are possible, his theory of fairness represents an attempt to formally model what structures and institutions are actually chosen.

But even granting that Binmore’s first two arguments establish the set of feasible, efficient outcomes, his theory of fairness does not provide the needed selection mechanism—both because his naturalized Rawls demands too much and explains too little. It posits the evolution of a relatively homogenous conception of fairness; yet the concept of fairness varies across and even within cultures. Binmore argues that our sense of fairness developed as an evolutionary byproduct of repeated games our ancestors confronted when trying to solve simple coordination problems. This appears problematic, since even the most basic fairness norms that govern family life differ widely from one culture to the next. The familial culture in Japan, for instance, is structured much differently than most of the world and it appears clear that these differing notions of fairness within a family can affect a society’s political conception of fairness. Modern philosophers even have a difficult time deciding what resource conceptions of equality are meant to address (Sen 1979). Even more problematic is the conflict between desert and equality. Should a universal healthcare system provide free help for everyone, even those (like freeform climbers) who willingly risk their lives knowing the risk? Or should equality-promoting institutions only focus on helping those who are less fortunate through no fault of their own? Different societies have come up with different answers to these questions, challenging Binmore’s notion of a generic evolved Rawlsian fairness. The empathetic mindset characterized by the Rawlsian system is certainly a part of our evolved morality but it would be too optimistic to think it is the primary equilibrium-selecting device.

To recap, Binmore’s notions of both the feasible and efficient institutions is theoretically clear and useful. These conditions lead to important practical and theoretical results. Further, his theory provides ample check against those who seek to produce ‘utopian’ social change. We are limited by the incentives facing agents and the constant potential gains from social defection. But unlike early scholars
who thought that such an atomistic decision theory framework was doomed to lead to strife and insurmountable collective action problems, Binmore’s development of the folk theorem for repeated games gives a strong theoretical framework for the empirical fact that people manage to overcome serious coordination problems without outside enforcement. Yet, as Binmore recognizes, the folk theorem and its game theoretic derivatives are not enough for a full theory of institutional development—they merely delineate the set of all feasible and competitive social contracts. They do not provide a means of selecting between them. Binmore’s own solution to this selection problem—the evolution of fairness norms—relies too narrowly on the Rawlsian interpretation to provide a comprehensive explanation of how specific institutional arrangements are chosen. A different equilibrium selection device appears to be required.

3. Storr and the Ideological Entrepreneur

Having examined both the analytical successes and potential problems with Binmore’s model, the challenge is to retain the formal game theoretic constraints used by Binmore while replacing his particular theory of equilibrium selection with a broader approach. One potentially promising way to approach this challenge is through the lens of Virgil Storr’s ideological entrepreneur. Storr (2009) presents his entrepreneur as a friendly challenge to Douglass North’s theory of institutions, particularly a tension he identifies between North’s theory of institutional path dependence and his brief discussion of the role ideological entrepreneurs have in setting the baseline for formal institutional cooperation. North argues that past institutional arrangements have a direct, and in many cases potentially insurmountable, effect on current societal norms. As Storr notes, North’s theory of path-dependence is, “not simply a claim that the past affects the choices … in the present[,] it is a description of how the dead hands of the past reach up from the grave to constrain and direct the living” (ibid.: 110). But, as Storr observes, swift institutional change is not only possible, it occurs frequently throughout human history. He writes at length about particular positive institutional changes in the Bahamas that happened over the course of a few short years. The horrifying institutional change that swept through Nazi Germany offers another example (ibid.). Swift changes in the formal institutional structure require at the same time a rapid shift in the underlying ideological climate, without which new formal institutions

7 Also see Storr 2004.
will not be supported by the informal social contract. Swift changes such as these cannot be accounted for in a heavily path-dependent world—they represent a movement away from the status quo due to deliberate human action and purpose. Storr argues that human action directed at shaping the ideological beliefs of fellow citizens is broadly similar to the key role the Kirtznerian market entrepreneur plays in the Austrian economic theory of the market process.

Having motivated the need for an Austrian entrepreneur, Storr gives a sketch of what both a Schumpeterian and Kirznerian ideological entrepreneur would look like in a theoretical model. Unlike North’s brief discussion of ideological entrepreneurship, an Austrian ideological entrepreneur would not be a neo-classical maximizer of a given choice set inherited from the past. These new dynamic, creative, and perceptive entrepreneurs would seek ideological “profit” by capturing the populace’s demand for new ideas—they would be alert to opportunities for “institutional arbitrage.” Further, they would merge and combine existing ideologies in order to compete more efficiently in the ideological marketplace—they would be “creative destroyers” of existing ideology (ibid.).

Storr’s theory and use of an Austrian ideological entrepreneurship is an excellent first step toward answering the question of how institutional change unfolds. But it raises as many questions as it answers. While there do appear to be cases of rapid institutional change, there also appear to be many situations in which, presumably, ideological “profit” could be made, yet no change has occurred. For every story of swift institutional change there are many institutions, both good and bad, that persist despite all efforts to the contrary. North’s path-dependence thesis seems to outweigh any claims of institutional entrepreneurship in these situations. A fuller theory of Austrian ideological entrepreneurship needs to identity not only what functions and roles an ideological entrepreneur fills in the story of institutional development, but model his constraints. Clearly such constraints will be more complicated than the ones faced by market entrepreneurs. When judging the result of the market entrepreneur, we have the useful metric of monetary profit to determine the inter-subjective value the community attributes to a good. No such immediate feedback is available in the case of the ideological entrepreneur—thus we have no way of modeling when such a figure’s actions will lead to a socially beneficial outcome, a regressive outcome, or have no effect. We need a better theory of constraints to explain why in some cases Austrian ideological entrepreneurship can help spur institutional development, and in others cases, North’s path-dependence rules the day.
In some ways then, an Austrian theory of ideological entrepreneurship has the opposite challenge that Binmore’s game theoretic account of the development of institutions faces. While Binmore has a robust method of modeling the necessary and sufficient conditions for stable and efficient social contracts, I have argued his particular equilibrium selection device is underdetermined. On the other hand, a Kirznerian theory of ‘social-contract’ entrepreneurship provides a good explanation of the mechanism by which actors have incentive to push for change and how that change is actualized. Yet as I have argued, it is unclear why we see rapid institutional change is some cases and what appears to be complete path dependency in others. Clearly, ideological entrepreneurs can, and in some cases do, break path dependency, thus leading to not only a change in informal institutions but the formal institutions which they support. The question is why this does not happen all the time.

We can deploy the efficiency and stability criteria presented in Section 2 to make some progress towards this goal. Ideological entrepreneurship in the Austrian sense is always occurring, but it will only have an effect on the populace at large if it is both presented in a favorable environment and appeals to a possible and plausible future outcome. In these situations the Austrian ideological entrepreneur will provide a good explanation of the development of informal, and thus indirectly, formal institutions. But in cases in which the ideological entrepreneur attempts to challenge a long-held institutional belief without a sufficiently compelling and plausible alternative, North’s path-dependence correctly predicts that the choices of the past strictly constrain the institutional choices of the present. We can characterize the possibility of a given ideological entrepreneur being successful on two fronts:

(1) In general, an ideological entrepreneur who is acting in a period of relative instability has an advantage over the reformer who is trying to change a relatively stable institution. To put this insight in game theoretic terms, a stable situation is one in which no player has an incentive to defect from cooperation. The fact that situations are often not stable (in the Nash equilibrium sense) allows for the possibility of ideological entrepreneurship to be effective in fostering institutional change. But we can certainly imagine scenarios in which institutional arrangements are stable, e.g., a convention of driving on the left side of the road. In these cases, it will be very difficult for an ideological entrepreneur to enact change. All institutional change has a cost—and if no one has an incentive to defect from
the current institutional environment, it will be very difficult to persuade others to adopt a new institutional arrangement.

(2) The choice-set available to an ideological entrepreneur may not be limited strictly by the past as North’s path-dependence suggests, but nor is the choice set open. Ideological alertness and innovation can only be profitable if it coheres with a given mindset of the population—ideological entrepreneurs do not invent a completely new way of looking at the world, they specifically target and enhance certain given beliefs and goals that already exist in the populace (Lavoie and Chamlee-Wright 2000). Coherence between a socio-cultural context and an ideological innovation reduces the cost of ideological reform while allowing agents to see how the alternative future promoted by the ideological entrepreneur might be stable, i.e., one in which no one has an incentive to unilaterally defect from cooperation. The game theoretic tools described in Section 1 provides the ideal toolkit for describing and analyzing a set of feasible equilibrium states—the efficient equilibria predicted by the folk theorem in repeated games.

4. A Possible Austrian Objection to the Folk Theorem

According to Kelly (2009), a primary motivation for the development of game theory was a reaction to Austrian ideas. Oskar Morgenstern, co-author of the foundational Theory of Games and Economic Behavior, (Morgenstern and von Neumann, 1944) was a student and follower of Mises. The only economics class John Nash ever took was from an Austrian professor. In his interview after his Nobel address he noted “by coincidence I was influenced by an Austrian economist which may have been a very good influence” (Nobelprize.org, 1994) But while the founders of game theory were certainly influenced by Austrian ideals, the methodological status of game theory within Austrian economics is unclear.

On the one hand, one might imagine the repeated game models suggested by the folk theorem might interest Austrians. These models establish the self-sustaining properties of evolved social institutions that depend on the incentives facing individual actors, not external enforcement. But as noted in the previous section, folk theorems do not rest upon Kirznerian alertness or Schumpeterian innovation; rather, they rest upon meta-strategies and conventional rules of social engagement. Further, there is no drive towards a particular equilibrium, just a calculation of whether a given situation is stable or not. But the interesting part of studying an economy is asking how social cooperation is achieved, not merely stating whether such states are stable.
As Boettke et al. (2003: 7) observe, when theorists rely completely on the folk theorem, “it is far from clear [the actors] would be able to obtain an equilibrium given the constant introduction of new knowledge and technology.” Further, they argue that there is something problematic about the non-universality of equilibria predicted by the folk theorem, being that “they happen to hold at a particular time and place … not necessarily hold[ing] in all cases with similar circumstances” (ibid.: 7). Such a multitude of equilibria leads to what they term “formalistic historicism,” in which economic analysis is aimed at retroactively justifying any particular states of the world rather than producing universal propositions. In short, they argue that because of the nature of game theoretic assumptions and the multiplicity of equilibria, folk theoretic analysis can fail to explain the dynamic action of man. This critique is by no means endemic within the Austrian community, and it is certainly possible that some who consider themselves Austrians may not find this a convincing critique. But given that the criticism strikes at the heart of what many would argue is the central aim of economic inquiry, it deserves a response.

While it is certainly true that the folk theorem merely states static possible states of the world, it would be unwise to dismiss the theory on methodological grounds alone. Combined with the right tools, it provides a positive compliment to Austrian ideas. The traditional Austrian conception of entrepreneurship is constraint-less, in the sense that a potential entrepreneur’s alertness to opportunities is not bounded by society. For market entrepreneurship, this assumption is not problematic since the ex-post fruits of successful entrepreneurship are directly observable in the profits and losses. No such clear signal exists for ideological entrepreneurship. A market entrepreneur exits if his venture fails but since ideological entrepreneurship happens outside the traditional market, we cannot rely on traditional economic wisdom to predict when a given act of ideological entrepreneurship will fail or succeed. There might be many ways to surmount this problem, but an appeal to the folk theorem and the robust collection of results tangent to the folk theorem in game theory provides a promising approach. The Austrian critique of the folk theorem assumes that most of its users are content to model formal conditions of cooperation without doing the dirty work and examining which human action actually produces cooperation. While that was true of many game theorists in the past, the current multidisciplinary program that uses the folk theorem and game theory more generally to study the evolution of norms has explicitly incorporated theories of action that defuse these critiques. Binmore’s
Rawlsian move is one flawed example. Replacing these problematic theories of action with ideological entrepreneurship strengthens both theories: Game theorists get a plausible selection mechanism and Austrian ideological entrepreneurship theories get a ready-made theory of constraints and feasibility conditions.

5. Conclusion

In this paper I have laid the groundwork for a theory exploring how entrepreneurs shape institutions. I focused on the construction of informal institutions—what I call social contracts or ideologies—on the assumption that before we understand what causes the development of particular formal institutions, we need to understand how the informal institutions that underlie them develop and change. I considered two theories, broadly similar in scope and goals yet methodologically very different, that attempt to explain how informal institutions develop: the game theoretic model of social contract evolution presented by Binmore, and the Austrian ideological entrepreneur as presented by Storr. Each theory was found to be promising but incomplete, and I have argued that a hybrid theory that combines the best features of each would provide a promising theoretical springboard for other scholars interested in the development of cooperative institutions.

References


