When Order Affects Performance: Institutional Sequencing, Cultural Sway, and Behavioral Path Dependence

Jenna Bednar and Scott E Page*

June 2, 2014

Abstract

How does the order that laws and other institutions are introduced affect their performance, and under what conditions will sequencing matter? To gain a foothold on these questions, we construct a formal model of institutional sequencing that includes cultural sway in the form of behavioral spillovers. We derive two broad categories of results: First, we characterize the relationships between the extent of cultural sway, payoff structures, and path dependence. We find that path dependence increases and then decreases in cultural sway and that optimal payoff structures in novel games require weak punishment regimes. Second, we characterize the optimal sequencing of institutions. We show that optimal sequences satisfy a property we call *multi-incrementalism* that combines initial institutional diversity with gradual movements to reduce inefficient spillovers. Such sequences have the counterintuitive property that they enable the *potential* for path dependence in order to avoid its *realization*.

Keywords: *Institutional performance, gradualism, transitions, learning, path dependence, equilibrium selection, quasi-parameters*

*University of Michigan, `jbednar@umich.edu` or `scottepage@gmail.com`. The authors gratefully acknowledge the hospitality of INSEAD during a sabbatical leave and research support from the U.S. Army Research Laboratory and the U.S. Army Research Office under contract/grant number W911NF1010379.

Introduction

Institutions are the tools humans use to achieve common goals, but their effect often strays from aspirations.[1] By designing and implementing formal institutions—rules and laws—humans intentionally create environments that facilitate coordinated and cooperative behavior in political, social, and economic activities. The planned elements of incentive environments interact with broader social forces, including beliefs and norms, at times generating unintended or unanticipated consequences. In particular, while each institution operates within a well defined domain—markets allocate goods, governments pass laws, and schools educate children—they also interact with one another. Many interact directly, as is the case with separation of powers or levels of government in federal systems.[2] Institutions also affect one another *indirectly* by creating belief systems and behavioral repertoires that can influence how people respond in other institutional settings. In this paper, we investigate the implications of these indirect effects on institutional performance and the optimal sequencing of institutional designs.

Our analytic approach will be to model individuals as having *behavioral repertoires*, a collection of behaviors that individuals apply in different contexts. These behavioral repertoires comprise a part of culture—a much larger bundle that includes beliefs, traits, symbols, and meanings (Swidler 1986). Modeling individuals as bundles or vectors of traits has a long tradition (Axelrod 1997, Boyd and Richerson 2005, Bednar and Page 2007). Here, we make two significant innovations. First, we construct the relevant repertoire sequentially by adding behaviors over time. Second, we assume the possibility of behavioral spillovers: when an

[1] Greif 2006, chapter 2, describes well the tension between the view of institutions as a product of human agency and as systems that transcend human design, and shares our view that institutional analysis should take both perspectives into account.

[2] Many studies examine how multiple institutions co-constrain or jointly motivate a particular choice. See for example, Putnam's (1988) two-level games or Tsebelis' nested games (1990) as well as Tsebelis 2002 and Weingast 1998 in which multiple institutions serve as constraints. These direct interactions can enable an assembly of imperfect institutions to improve upon the capacity of any one institution acting on its own (Bednar 2009, Vermeule 2011).

individual encounters a new institution, with some probability their initial response will be drawn from their existing repertoire.

Our central premise is that because individuals interact across multiple institutional settings, the behaviors that emerge in any one context—be they cooperative, trusting, altruistic, or competitive—bleed into other institutional settings, creating consistency of behavior across contexts.[3] These behavioral spillovers in turn shape individual behavior, and, through aggregation, social outcomes. Thus, the spillovers underpin a logic for how existing institutions can either facilitate or undermine the success of new institutions by favoring certain behaviors.

Our formalism and our focus on behavior are novel but the core idea of institutional context mattering permeates the literature. In Long's (1958) conception of an "ecology of games", institutions create a behavioral or belief environment, and through that, affect the performance of other institutions. Similarly, Aoki's (1994, 2001) theory of complementary institutions assumes that the presence of one institution in an environment makes another more effective, and North's (1993) "institutional matrix" (along with beliefs) causes institutional change to be incremental. Finally, Mahoney and Thelen (2010) include the role of agency and a dynamic environment in modeling incremental institutional change. Each of these models includes spillovers, but none rely on behavior as the fundamental unit of analysis.[4] We believe that a model built upon behavior has several advantages. Most no-

[3]These behavioral spillovers are the core assumption of our model. Empirical support for their existence can be found in multiple disciplines using diverse methodologies. Fieldwork by social psychologists shows that routine actions can shape cognitive outlook (Talhelm et al 2014). In cognitive psychology, there exists a substantial literature on *cased based* reasoning (see Gilboa and Schmeidler (1995) for a summary) as well as an extensive literature on cultural priming by cultural psychologists. For example, experiments demonstrate the ability to prime individualist and collectivist behavior, showing that behaviors respond to cultural cues and are not static (see Oyserman and Lee (2008) for a meta analysis). Anthropologists and economists have run common experiments in distinct cultural groups and found that responses align with cultural practices (Henrich et al (2001, 2004). And finally, work by experimental economists on multiple game experiments find support for cross-game spillovers (Bednar et al 2012, Cason et al 2012). At a more macro level, the assumption of spillovers producing consistency also aligns with cross-national survey research on cultural diversity (Inglehart 1990, 1997).

[4]One could also motivate an assumption of diverse behaviors by referring back to heterogeneous beliefs.

tably, it can be measured in the field and in the laboratory, which is why there exists the aforementioned evidence of spillovers from various disciplines.

As mentioned, we assume that institutions, which we represent as games, are added sequentially and the the repertoires accumulate. How individuals behave initially in a new and novel game depends on the extent of what we refer to as *cultural sway.* Those individuals subject to cultural sway choose an initial strategy that matches the equilibrium strategy in the most similar existing game. In effect, they think "this new game looks a lot like this other game that I already play, so in the new game I'll initially try the behavior I use in the other game." Other individuals approach the new game with a blank slate. They view it detached from any cultural forces and focus entirely on the payoff structure. This leads them to choose the payoff maximizing equilibrium action initially.

These two types of individuals—those whose behavior is culturally embedded and those whose initial behavior is context-free—then interact within the institution. They apply a rational learning rule, the outcome of which depends on the distribution of initial behaviors. That equilibrium could be the culturally influenced behavior, or it could be the efficient outcome. The behavior that emerges depends on the extent of the cultural sway as well as the payoff structure to the game. Once established, the behavior gets added to the repertoire. It becomes part of the culture. Note that culture does not determine outcomes, but that initial behaviors for some people are influenced by culture. These initial behaviors can in turn influence the outcome in some circumstances. The model provides a framework for characterizing the conditions that generate path dependent—ie suboptimal—behavior, and the generation of an institutional sequence that is most likely to lead a society toward its goals.

Although abstract, the model's construction was motivated by explicit policy domains and questions raised by scholars and policy makers interested in institutional remedies to complex problems, where sequencing of institutions is necessary. The notion that the performance of an institution depends on pre-existing institutions, and the order that those institutions were introduced, is broadly relevant. The reconstruction of the United States after the Civil War, the introduction of elections in nascent democracies, or reduction of greenhouse gas emissions are each multidimensional policy objectives, requiring a set of institutions to accomplish, and therefore requiring a determination of the order that each should be introduced to maximize the likelihood that any particular institution is successful and that society comes closest to achieving its goal. Interactions between institutions imply that some sequences will produce more efficient outcomes than others, that some sequences are more likely than others, and that behavior and institutions change in tandem and do so coherently.[5] When more than one institution is needed, a necessary question is how the institutions should be rolled out: simultaneously or sequentially, and if the latter, in what order to make each institution most effective? Our model suggests that early institutions produce behavior that can contribute to or undermine the success of later institutions. If that's true, then those who design those transitions could be helped by an understanding of how, when, and why sequencing matters.

One multi-dimensional problem space that occupies the attention of many scholars is development. Theories of democratization and economic development imply conflicting sequencing prescriptions. Most generally, there is the competing approaches to timing, characterized as *gradualism* (eg. Dewatripont and Roland 1992, Carothers 2007, Roland 2000, 2002) vs. *big bang* (Lipton and Sachs 1990). Big bang, or shock therapy, advocates radical and comprehensive (multi-institutional) departures from existing institutions for quick improvement, while with gradualism, steps are taken toward the social goal that begin from the baseline of existing conditions, working with the positive aspects of a political economy, rather than strictly against the undesirable aspects. New institutions are introduced slowly and start with reforms considered most likely to be popular or successful (Roland 2000), as

[5] See Inglehart (forthcoming).

public acceptance for reform builds.

Other sequencing arguments are more qualitative, prioritizing democracy (Sen 1999, Carothers 2007, Berman 2007, Knight and Johnson 2011), or growth and its enabling institutions (Lipset 1959, North and Thomas 1973), or civil society (Huntington 1968, Putnam 1993), or security (Mansfield and Snyder 2005, Lake 2010). For all this theoretical diversity—and there are dozens more that we could mention on this topic—the studies are united by two traits: they are excellent scholars doing careful work, and they think that order affects outcomes. They just disagree about the optimal ordering.

Formal analysis of the influence of sequencing on institutional performance means situating an institution within a context of institutional and behavioral space. In historical institutionalism scholars focus explicitly on space and time by examining context to understand how experience shapes responses at particular moments. In the methodology of process tracing, scholars identify key causal mechanisms within the context of a specific case (Thelen 1999, Mahoney 2001, 2010, Brady & Collier 2004, Falleti & Lynch 2009). As Falleti and Lynch, quoting Goertz (1994), put it: "Context plays a radically different role than that played by cause and effect; context does not cause X or Y but affects how they interact" (Falleti & Lynch 2009:1151; Goertz 1994:28).

Identification of preconditions, or drawing on case studies and experience, provides an evidence-based foundation for planning the sequencing of institutions. A complementary approach is to construct a model based upon a theory of how institutions relate to one another. That is the route we pursue here. The question of how institutions establish a context that affects their own performance as well as the performance and choice of future institutions has been examined by scholars interested in transitions to democracy and market based economies (e.g. Roland 2000) and in studies of endogenous institutional change (Greif and Laitin 2004, Greif 2006, Mahoney and Thelen 2010). Historical narratives situate institutions' contextual effects in beliefs, behaviors, norms, rituals, habits, and organizations

(Grief 2006), but any formal model must reduce this dimensionality of causes. Grief, for example, relies on *beliefs* as the cultural attribute that transmits the weight of past institutions and constrains the set of equilibria. Beliefs, in turn, can determine public acceptance of institutions (Roland 2000). We pursue a different analytic path and focus on how institutions function in light of preexisting behaviors, which carry cultural tendencies from one institution to another.

Our behavioral approach does not contradict belief-based models, but does enable us to explore a different set of questions and to draw distinct insights. Both model types identify conditions for institutional path dependence. Belief-based models require constraints on priors, while our model requires lower bounds on the extent of the cultural sway. As a result, the models come to different normative conclusions about institutional choice. For example, Greif (2006) highlights a fundamental asymmetry between institutions that build from existing structures and those that are created de novo. He derives a strong preference for the former because the latter lack sufficient context for similarities in beliefs. As a result, learning will be a "lengthy, costly, uncertain endeavor." Greif concludes that human nature advantages traveling familiar paths. A society's historical experience with an institution, or components of it, should cause that society to implement familiar institutional components rather than ones that might appear to be more efficient, from a mechanism design perspective. There's efficiency in familiarity.

Greif's prescription matches the gradualism approach from the development literature. Our model provides a basis for evaluating the prescription of gradualism. With the model, we can isolate the conditions when gradualism would and would not lead to optimization of social goals.[6] As we show, building on a single behavior generically leads to inefficient outcomes, suggesting that there are limits to the appropriateness of gradualism, or one-

[6]As should be clear from our framing, the point of our model is not so much to derive testable predictions or to fit history exactly, but, following Johnson (2014), to uncover the core logics.

sided incrementalism. We find instead that optimal sequencing requires diverse institutional forms early (strong incentives for new behaviors *and* weak punishments for experimentation) followed by *multi*-incrementalism—minor adjustments from institutions that produce diverse behaviors. Early diversity builds dimensionality to behavioral repertoires, resulting in greater responsiveness to incentive structures, for example enabling people to behave competitively in a market setting and cooperatively within their communal organizations. This diversity allows for potential path dependence, but importantly, that dependence won't be realized. What will be realized are efficient outcomes through the incremental changes. This incrementalism creates the potential for path dependence but also prevents it from occurring.

We derive five results related to cultural sway and institutional performance and three results on optimal sequencing. First, we find that cultural sway can produce suboptimal equilibrium strategies. Second, we demonstrate that any set of institutions will be subject to some degree of behavioral path dependence unless those institutions all have unique equilibria. Third, we relate the amount of cultural sway and the set of previous games to the extent of path dependence. Fourth, we show as cultural sway becomes large path dependence can decrease and initial institutional choice can determine outcomes. Finally, we show in a general class of games that optimal institutional design implies strong incentives to choose an equilibrium but weak punishments for deviating.

The first of our three results on optimal sequencing states that the most efficient paths—when agents maximize their payoffs—include more diverse games earlier in the sequence and then relies on incrementalism. Second, we show that this optimal sequences enables the possibility of path dependence but then avoids it. Third, borrowing the quasi-parameter model of Greif and Laitin (2004), we find that negatively reinforced institutional drift leads to institutional change at an inefficient moment.

We've organized this paper into five parts. We first present our modeling framework. We then apply our model to two families of games that can be parameterized by a single variable and include a sketch of how one might apply the model to sequencing in democratic transitions. We then present our main results. In the fourth part, we extend the model to cover a general class of all two by two symmetric games as well as arbitrary game forms. We conclude by relating our results to the literature and discussing possible extensions.

1 The Model

We construct a model with the following three characteristics: institutions arrive sequentially, individuals have different priors about how to respond to a new institution, with some drawing from past behaviors and others immediately playing the payoff maximizing strategy, and in subsequent periods individuals learn their way to an equilibrium in the new institution.

We consider an infinite population of individuals who play a sequence of games with one another. We divide our population into two categories: those whose initial action is embedded in *culturally embedded* current practices and those who are *context free*.[7] The former group sees the new game within the context of existing institutions and chooses a familiar behavior, the equilibrium strategy employed in the closest game, perhaps because it reduces cognitive costs or because people naturally reason by analogy.[8] Those who reason *context-free* approach the game with a blank slate. They interpret the game devoid of any context, in the same way that someone trained in game theory might look at a payoff matrix in an experimental setting. Their initial action in the game is that which produces the highest payoff if all individuals take that action.

We divide time into two components: *epochs* and *periods*, where each epoch is divided into a large number of periods. In each epoch we introduce a new game. That game played some large finite number of periods within the epoch. We remain agnostic as to whether that same game is played in subsequent epochs. If it is played in future epochs, we assume

[7] In experimental settings, initial play in games is typically heterogenous (Camerer 2003).

[8] See Samuleson (2001), Gilboa and Schmeidler (1995), Jehiel (2005), and Bednar and Page (2007).

that individuals continue playing the same strategies.

Each game is chosen from a family of symmetric games, G. We denote the game selected in epoch t by g_t. We denote the payoff maximizing repeated game strategy by s_t^*.[9] In the first epoch, we assume that all individual choose s_1^*, the focal, payoff maximizing equilibrium strategy. In all subsequent epochs, we assume that individuals choose initial strategies according to their types as described above. After the initial period, the population learns an equilibrium behavior using best response learning (Nash 1951).

The probability that an individual's initial behavior is culturally embedded or context free depends on the amount of *cultural sway*. We formalize this with a parameter γ, which equals the probability that an individual's initial response draws on a preexisting behavior. Formally, we assume that a proportion γ of the individuals compare the new institution with all existing institutions (games), identifying the game in the sequence that most closely resembles game g_t, and play that strategy in the new game. The remaining fraction $(1-\gamma)$ of the individuals whose behavior is context free choose the payoff maximizing equilibrium strategy, s_t^*.

Identifying the nearest prior institution requires a distance function between games. In our first two sets of models, games will be identified by real-valued parameters. We use the Euclidean distance between payoffs as a distance measure. In the one-dimensional model presented in the next section, we index games by θ, a real-valued parameter. The distance between two games equals the absolute value of the differences between their θ's. Later, we consider more general distance functions.

We define an *historical context* as a spillover parameter together with a sequence of games $\Omega = \{\gamma, (g_1, g_2, ..g_k)\}$. For notational convenience, we denote the equilibrium outcome in game g in context Ω by $\Omega(g)$. We define a *continuation* to be any sequence of games adde to an historical context. Two historical contexts are *next outcome equivalent* if given any single

[9]In the event that there exist multiple payoff maximizing strategies, we assume that one is focal.

game continuation they produce the same outcome.

Contexts $\Omega = \{\gamma, (g_1, g_2, g_3, ..g_k)\}$ *and* $\hat{\Omega} = \{\hat{\gamma}, (\hat{g}_1, \hat{g}_2, \hat{g}_3, ..\hat{g}_{k'})\}$ *are* **next outcome equivalent** *if and only if given any game* $g \in G$, $\Omega(g) = \hat{\Omega}(g)$

One of our main areas of interest is in how different contexts enable path dependence. A set of games $\{g_1, g_2, g_3, ..g_T\}$ produces **path dependent behavior** if two distinct reorderings of those games result in different equilibrium play in an least one of the games (see Page 2006). We are interested comparing levels of path dependence and comparing across sequence-built contexts. To do so requires additional definitions. Define a *continuation path of length* m, C_m, to be a sequence of m games and let $\Psi_m = \{C_m : C_m = (g_1, g_2, g_m)\}$ denote the set of all continuations of length m. Define a *context and path* as a context together with a continuation path of length m. Denote this by (Ω, C_m).

Given two next outcome equivalent contexts, Ω *and* $\hat{\Omega}$*, context* Ω *exhibits* **greater path dependence** *than context* $\hat{\Omega}$ *if and only if either of the following two equivalent conditions hold.*[10]

(i) For any game, the set of continuation paths that changes the outcome in game g in context Ω *strictly contains the set of continuation paths that change the outcome in context* $\hat{\Omega}$.

$$\left\{C_m \in \Psi_m : \hat{\Omega}(g) \neq (\hat{\Omega}, C_m)(g)\right\} \subset \{C_m \in \Psi_m : \Omega(g) \neq (\Omega, C_m)(g)\} \;\; \forall g \in G \; \forall m \geq 1$$

(ii) For any game along any continuation path, the number of times the outcome in g changes in context Ω *is greater than or equal to the number of times it changes in context* $\hat{\Omega}$*, with a strict inequality for some paths.*

[10] To prove that these definitions are equivalent is straightforward. Suppose that the first condition does not hold. Then the continuation path for which a game changes under $\hat{\Omega}$ and not Ω contradicts the second condition. If the second condition is violated, then choose a continuation path under which a game switches outcomes more often in context $\hat{\Omega}$.

The performance of some institutions depends on the order that the institutions were introduced prior to its appearance. We refer to these institutions as *susceptible*: their outcomes can change along a continuation. If a game's outcome is not a function of the sequential context, we refer to it as *immune*. Finally, in contexts with substantial cultural sway, the first game can have a large effect on future outcomes. We define the *extent of initial game dependence* for a context Ω to be the probability that the outcome of a game in the susceptible region will be the same as that of the initial game in the context.

2 Games of Tradition and Trust

To build an intuition for the primary concepts in our model, we derive results for two familiar classes of games that can be parameterized along a single dimension. The first family of games considers situations in which individuals can choose a *traditional* action or an *innovative* action. The second family of games considers situations in which individuals can either take a safe action or a trusting action.

Upholding Tradition or Being Innovative

In the first family of games, individuals must decide to whether to stick to tradition or to adopt an innovation. The payoffs to each action are determined by a parameter $\theta \in [0, 16]$. If both players stick to tradition, each gets a payoff of $(16 - \theta)$. If both play an innovative new action, each gets a payoff of θ. There exist two pure strategy equilibria for $\theta \in [4, 12]$: sticking to tradition (T) and innovating (I). Sticking to tradition will be efficient if $\theta \leq 8$ and innovating if $\theta \geq 8$. If the two players choose opposite actions then each receives a payoff of zero.

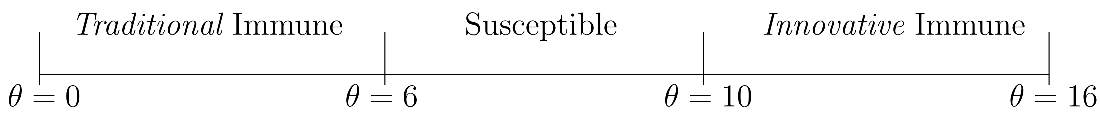

Figure 1: Susceptible and Immune Regions in the Tradition/Innovative Game as a Function of θ

	Traditional (T)	Innovate (I)
Traditional (T)	$16-\theta, 16-\theta$	$4, 4$
Innovate (I)	$4, 4$	θ, θ

Let the amount of cultural sway, γ, equal $\frac{3}{4}$, so that three-fourths of the population plays the equilibrium action from the closest game. Assume that the first game in a sequence of games has $\theta_1 = 7$. By assumption, the outcome in the first game will be efficient, so individuals will choose the traditional strategy. Assume that for the second game, $\theta_2 = 9$. Initially, three-fourths of the population will play the traditional strategy and one-fourth choose to be innovative. The payoffs for the two strategies in the population are as follows:

Traditional (T): $\frac{3}{4}(7) + \frac{1}{4}(4) \ = \frac{25}{4}$

Innovative (I): $\frac{3}{4}(4) + \frac{1}{4}(9) \ = \frac{21}{4}$

For $\theta = 9$, if everyone were to be innovative they would earn higher payoffs, but the payoff to being traditional is higher given the amount of culture sway. If in subsequent periods people learn to play the strategy with the higher payoff, then the traditional strategy will come to dominate. Thus, in the learned equilibrium everyone chooses the traditional action.

If instead, the first game in the sequence had produced innovative strategies, i.e. that $\theta_1 > 8$, the outcome in the second game with $\theta_2 = 9$ would also have been innovative. Given that the outcome in the game $\theta_2 = 9$ depends on the games that precede it, it is *susceptible*, a condition that is required for a game to have a path dependent outcome. In this example, the sequence of games $(\theta_1 = 9, \theta_2 = 7)$ produces innovative outcomes in both games, where as we just showed, the sequence ($\theta_1 = 7, \theta_2 = 9$) produces traditional outcomes in both games. Hence, outcomes depend not just on the set of games, i.e. *set dependence* (Page 2006), but also on the order in which those games are played.

Not all games will be susceptible. If θ is sufficiently high (resp. low) then the outcome will be innovative (resp. traditional) regardless of the previous games, as depicted in Figure 1. To see why, suppose that the first game in a sequence produces an efficient, traditional outcome, e.g. $\theta_1 < 8$. If the second game has $\theta_2 > 10$, then both players choose innovative actions despite cultural sway.[11] A similar calculation shows that for $\theta_t < 6$, the strategy chosen will be traditional regardless of the previous games played. Therefore, the values $\theta = 6$ and $\theta = 10$ partition the parameters into the *immune* and *susceptible* regions.

Extension: Sequencing of Electoral Institutions

The coordination game considers only two players. The model can be extended to cover games with arbitrary numbers of players. For example, we could apply the coordination model to the study of electoral behavior. Consider voters who must coordinate on either regional or national parties in a series of two elections: one regional and one national. If regional elections are held first, then voters would be more likely to coordinate on regional parties. When subsequent national elections are held, they may also support regional, as opposed to national parties. In contrast, if the national elections were held first, then national parties would be more likely to emerge and then when subsequent regional elections are held, those nationalist behaviors would spill over into the regional election. The relevant behaviors here would encompass more than voting and include gathering information, developing policy platforms, and forming relationships with people outside the region. These might all transfer

[11]The payoff to the traditional action equals $\frac{3}{4}(16 - \theta_2) + \frac{1}{4}(4) = 13 - \frac{3}{4}\theta_2$. The payoff to the innovative action equals $\frac{3}{4}(4) + \frac{1}{4}(\theta_2) = 3 + \frac{1}{4}\theta_2$. The latter exceeds the former if and only if $\theta_2 \geq 10$.

to the other elections.[12]

Linz and Stepan (1992,1996) make these same arguments (without relying on a formal model) to argue that transitions to democracy should hold national elections first. As a success story, they cite Spain, where national parties won a majority of the vote in early elections, despite strong Basque and Catalan regional identities. They contrast this with the processes carried out Yugoslavia and the former Soviet Union in which regional elections where held first. Our model provides a simple framework for describing their logic. It also identifies the conditionality of their claims. Applying our model, it follows that if the payoffs from coordinating on regional interests in say Moldavia, Georgia, and Ukraine, were sufficiently strong, then regionalist behavior could have lain within the *immune region.* If so, even if voters had voted for national parties in the national elections, voters would have coordinated on regional parties in the regional elections. Linz and Stepan's argument that holding national elections first would have solved the problem implies a bound on the level of attachment to regional identities sufficient to keep the regional elections in the susceptible region. In specifying conditions, our model helps to build a testable hypothesis of their argument.

[12] A formal version might look as follows: assume N voters within a region who can choose to coordinate on *national* issues (U) or *regional* issues (R). Let N_R denote the number of people who regional issues and $N_U = (N - N_R)$, denote those who choose national issues. Using a crude variant of the cube rule (Taagepera and Shugart 1999), payoffs could be written as follows:

$$\pi_{REG} = \theta \left(\frac{N_R}{N}\right)^3 + (1-\theta)\left(\frac{N_U}{N}\right)^3$$

$$\pi_{NAT} = (1-\theta)\left(\frac{N_R}{N}\right)^3 + \theta\left(\frac{N_U}{N}\right)^3$$

where the parameter θ denotes the relative advantage of being regional focused in the regional election and nationally focused in the national election.

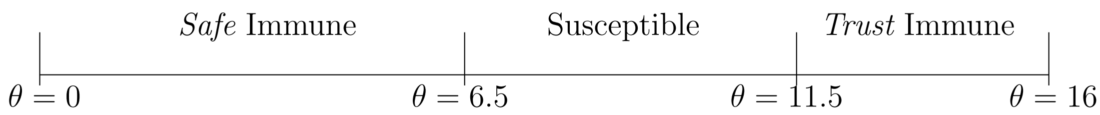

Figure 2: Susceptible and Immune Regions in Safe/Trusting Game as a Function of θ

Trust Games

We next consider a family of trust games in which there exists a *Safe* action and *Trusting* action. This family of games generalizes the familiar Stag Hunt game. Note from the payoffs that if $\theta \leq 8$ then Safe is the efficient equilibrium and if $\theta \geq 8$ both players choosing Trusting is efficient.

	Safe	Trusting
Safe	$16-\theta, 16-\theta$	$4, 2$
Trusting	$2, 4$	θ, θ

The initial susceptible regions are as shown in Figure 2.[13]

For this class of games, the immune regions are asymmetric favoring the Safe action. Safe has an advantage because it is *risk dominant* and generally speaking, learning advantages risk dominant institutions (Samuelson 1997). Trust, therefore, will be relatively harder to produce. Consider the following set of games $\{7, 9, 10, 11, 14\}$. If game $\theta = 7$ occurs first, then the only possible sequence that obtains the efficient outcomes in all remaining games is $(7, 14, 11, 10, 9)$. Notice that this sequence frontloads games in which Trusting will be the outcome and then incrementally builds trust.

[13] To solve for the boundary of the immune region for the trusting action, choose θ so that the trusting strategy receives a higher payoff even if three fourths of the individuals play the safe action. Formally, set θ so that $\frac{3}{4}(16-\theta_B)+\frac{1}{4}(4) \leq \frac{3}{4}(2)+\frac{1}{4}(\theta_B)$. Solving gives the threshold at $\theta = 11.5$. A similar calculation gives that the threshold for the safe action as $\theta = 6.5$.

3 Results: One-Dimensional Family of Games

We now state general results for a family of games that includes coordination games and trust games. Our results will rely on the concepts and intuitions developed in the examples. We make the following formal assumptions.

Assumption 1 *There exists a family of symmetric two by two games indexed by a one-dimensional real-valued parameter, $G(\theta)$, with $\theta \in [\theta_L, \theta_U]$ with two pure strategies denoted by A and B. Payoffs are maximized if both players choose the same strategy for all θ. Payoffs for A are maximized at θ_L and payoffs for B are maximized at θ_U.*

Assumption 2 *The payoff to playing B increases in θ and the payoff to playing A decreases in θ. These marginal effects increase in magnitude when the other individual chooses the same action.*[14]

Assumptions 1 and 2 imply that there exists an efficiency cutpoint, $\theta^{=}$, such that for any game $\theta \leq \theta^{=}$, A is payoff maximizing, and for any $\theta > \theta^{=}$, B is payoff maximizing. To simplify the presentation, we define $\theta^A(\gamma)$ and $\theta^B(\gamma)$ to denote the boundaries of the initial susceptible region. Thus, strategy A is immune for any game with $\theta < \theta^A(\gamma)$ and strategy B is immune for any game with $\theta > \theta^B(\gamma)$. If there exists no immune region for strategy A (resp. B) then we set $\theta^A = \theta_L$ (resp. $\theta_B = \theta_U$).

Our fist claim states that the size of the initial susceptible region increases in the size of the spillover, thus, the stronger the spillover, the likely inefficient equilibrium emerge in later games. The proof of this and all claims are in the appendix.

Claim 1. *Increasing the amount of cultural sway makes more games susceptible to sequencing: $\theta^A(\gamma)$ (resp. $\theta^B(\gamma)$) weakly decreases (increases) in γ.*

[14] Formally, this can be written as $\frac{\partial \pi_{BB}(\theta)}{\partial(\theta)} > \frac{\partial \pi_{BA}(\theta)}{\partial(\theta)}$ and $\frac{\pi_{AA}(\theta)}{\partial(\theta)} < \frac{\partial \pi_{AB}(\theta)}{\partial(\theta)}$, where $\pi_{ij}(\theta)$ equals the payoff to an individual playing i whose opponent plays j.

The next lemma that states that at the end of any sequence of games, there exists a threshold T such that in the next game, the strategy A will be played if $\theta < T$ and B if $\theta > T$. We refer to the equilibrium strategy that emerges from a game as the outcome.

Lemma 1. *The outcome in a game is determined by a threshold in the space of payoffs that depends on the sequence of past games and the amount of cultural sway. [In each epoch $t > 1$, there exists a threshold T_t such that if $\theta_t < T_t$, A will be the outcome and if $\theta_t > T_t$, B will be the outcome. T_t is a function of γ.]*

The threshold will equal the average of the largest θ that produces an outcome of A and the smallest θ that produces an outcome of B, provided that the average lies in the susceptible region. Therefore, it depends on both the spillover parameter and the payoffs in the first game. We now state a corollary that makes two points: first, the closer the first game is to the efficiency cutpoint the more it will affect later paths, and second, the greater the amount of cultural sway, the more the first game matters.

Corollary 1. *If the initial game produces outcome A, then for any sequence of games that follows, the threshold increases in the amount of cultural sway and in the payoff parameter θ_1. [Given $\Omega = \{\gamma, (\theta_1)\}$, where $\theta_1 < \theta^{=}$, for any continuation $(\theta_2, \theta_3, \ldots, \theta_k)$, the threshold at time k, T_k, weakly increases in both γ and θ_1.]*

Path Dependence and Initial Game Dependence

We now demonstrate how the extent of path dependence depends on parameters of the model and how levels of path dependence depend on the context in which a continuation occurs. We first state a sufficient condition for institutional path dependence to exist.

Claim 2. (Existence of Path Dependence) *Any set of games that contains at least one susceptible game and two games with different efficient equilibrium outcomes exhibits path dependence.*

The claim has a straightforward corollary.

Corollary 2. (Existence of Susceptible Games): *For any set of games that contains at least one susceptible game and two games with distinct efficient outcomes, for either outcome there exists an ordering of the games such that all susceptible games produce that outcome.*

Note that increasing the size of the susceptible region allows for the possibility of path dependence: what happens can depend on the games chosen. Susceptibility, though, differs from path dependence as shown in the next claim.

Claim 3. (Path Dependence need not imply Susceptibility) *There exist outcome equivalent contexts Ω and $\hat{\Omega}$ such that the susceptible region for Ω contains the susceptible region for $\hat{\Omega}$, but that context Ω does not exhibit greater path dependence.*

A larger susceptible region will imply greater path dependence if there has existed at least one outcome of each type in both contexts.

Claim 4. (Path Dependence Requires Diversity) *If two outcome equivalent contexts each include one outcome of each type, then a larger susceptible region implies greater path dependence.*

The previous claim implies that when deciding between two institutions with the same efficient equilibrium, choosing an institution with clearer incentives—i.e a θ further from the threshold—produces greater future path dependence.

Corollary 3. *Given any context, and two games that produce the same outcome in that context, choosing the game further from the threshold results in context with greater path dependence than choosing the game closer to the threshold.*

Our results imply that the degree of path dependence depends on cultural sway if both outcomes have occurred. However, in the limit as γ approaches one, the susceptible region can converge to the entire space. In those cases, the first strategy played first will be played in all games. This implies sensitivity to the initial game, but *not path dependence* The extent of initial game dependence can be measured as the probability that a future game has the same outcome as the first game. We next stat that the extent of initial game dependence strictly increases in the amount of cultural sway.

Claim 5. (Large Sway Implies Initial Game Dependence) *The extent of initial game dependence strictly increases in γ and approaches one as the amount of cultural sway approaches one.*

The previous claim describes outcomes for γ near one. The same effect hold for less cultural sway as well. In Figure 3, we show results from 1000 simulations of our Tradition/Innovation Game. We plot number of times that the final threshold was on the same side of the efficiency cutpoint as the initial game against the number of times that it did not. Think of this as the odds ratio for the initial game to have sway. For low amounts of cultural sway, the ratio is around two, which suggests path dependence. For large amounts of cultural sway, the odds ratio approaches seven; the initial game drives most of subsequent outcomes.

These numerical calculations demonstrate that if the outcomes depend on the path, then both action types are still in play. If outcomes do not depend on the path, and only depend on the initial game, then the system can get locked into one equilibrium—one behavior—for almost all games.

Efficient Paths

We now characterize efficient paths. We say that a sequencing is *efficient* if every game produces an efficient outcome. We first show that for some sets of games, no efficient sequence may exist.

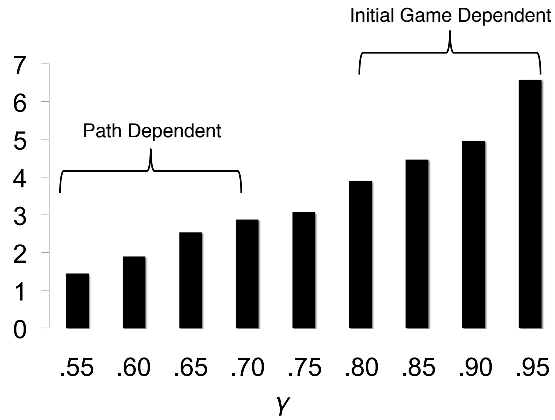

Figure 3: Odds Ratio of Threshold in direction of Initial Game in Tradition Game after 1000 Epochs.

Claim 6. (Optimal Sequencing Need Not Be Possible) *There exists sets of games such that for any sequencing of the games, the equilibrium selected will not be efficient in at least one game.*

We now derive a necessary and sufficient condition for an efficient path to exist. First, given any set of games $\{\theta_1, \theta_2, \ldots, \theta_N\}$, reorder the θ's from smallest to largest and relabel those θ's that are less than the efficiency cutpoint, $\theta^=$, by α_1 to α_R, where $\alpha_j < \alpha_{j+1}$. Next relabel the θ's that are greater than $\theta^=$ by β_1 through β_M where $\beta_i > \beta_{i+1}$. For a game α_j, define the *other equilibrium index*, $I(\alpha_j)$, as the number of games labelled β which are further from α_j than its distance from α_{j-1}. For an immune game, the index equals M. Suppose the α games are introduced in increasing order α_1, then α_2, and so on. It follows that *other equilibrium index* for the game α_j equals the maximal number of β games that can exist in the sequence prior α_j and still have α_j produce the efficient equilibrium. Note that the indexes need only be weakly increasing. The formal definition can be written as follows:

$$
\begin{aligned}
I(\alpha_j) \quad &= \max i \text{ s.t. } (\beta_i - \alpha_j) > (\alpha_j - \alpha_{j-1}) \text{ if } \alpha_j > \theta^A \\
&= \text{M} \quad \text{otherwise} \\
I(\beta_i) \quad &= \max j \text{ s.t. } (\beta_i - \alpha_j) > (\beta_{i-1} - \beta_i) \text{ if } \beta_i < \theta^B \\
&= \text{R} \quad \text{otherwise}
\end{aligned}
$$

Supposes that game α_j has an index larger than j, implying that games β_1 through β_j can occur ahead of it in the sequence and α_4 will still produce the efficient outcome. If this does not hold, if α_j has an index less than j, then game α_j (β_i) *exceeds balanced sequencing.* If no game exceeds balanced sequencing, then α and β games can be alternated ($\alpha_1, \beta_1, \alpha_2, \ldots$) to create an efficient sequence. But if game α_4 has an index of two, then it must appear prior to game β_3, so game β_3 must have an index of at least four. We can now state necessary and sufficient conditions for the existence of an efficient sequence.

Claim 7. (Sufficient Conditions for an Efficient Sequence) *Given a set of games $\{\alpha_1, \alpha_2, \ldots, \alpha_R, \beta_M, \beta_{M-1}, \ldots, \beta_1\}$, there exists a sequencing of the games that produces efficient outcomes in every game if and only if the following two conditions hold:*

(i) If $j > I(\alpha_j)$, then for any β_i s.t. $i > I(\alpha_j)$, $I(\beta_i) \geq j$.

(ii) If $i > I(\beta_i)$, then for any α_j s..t. $j > I(\beta_i)$, $I(\alpha_j) \geq i$.

The claim demonstrates that if an efficient sequencing exists, one way to produce it is to begin from the most diverse institutions (as determined by θ) and to work inward. Thus, societies that experience more diverse sets of institutions in earlier epochs may perform better.

That insight holds more generally; sequences that produce final thresholds near the efficiency cutpoint, $\theta^=$, are more likely to produce the payoff maximizing outcome. The following claim establishes a prescription for optimal ordering: Games with stronger incentives should be placed earlier in the sequence. We refer to this as *multi-incrementalism.*

Claim 8. (Multi-Incrementalism) *Given any set of games relabeled as* $\alpha_1 < \alpha_2 \cdots < \alpha_R < \theta^{=} < \beta_M < \beta_{M-1} \cdots < \beta_1$*, any game sequence in which there exists a* $j > j'$ *where* α_j *(resp.* β_j*) appears prior to an* $\alpha_{j'}$ *(resp.* $\beta_{j'}$*) produces inefficient outcomes in at least as many games as an alternative sequence in which game* α_j *appears before* $\alpha_{j'}$ *(resp* β_i *appears before* $\beta_{j'}$*).*

An example clarifies the logical argument. Let $\theta^{=} = 10$, $\theta^A = 6$, and $\theta^B = 14$. Consider the following sequence of θ's: (15, 8, 12, 7). In the first epoch, the outcome will be B and the new threshold will be 6. We write this as $(15B^*, 6)$; we place an asterisk to denote that the outcome is efficient. We obtain the following sequence of outcomes and thresholds: $(8B, 6)$, $(12B^*, 6)$, and $(7B, 6)$. Thus, two of the games produce inefficient outcomes. An optimal sequence would be (7, 15, 12, 8). This sequence establishes an A outcome before any B's; early diversity of games with strong incentives improves performance of other games in the sequence.

Many equate path dependency with *inefficiency*. But our model shows that choosing games that have clearer incentives—those further from the threshold—is optimal and increases *potential* path dependence. But by gradually introducing institutions from multiple directions, the *realization* of path dependence is minimized. Therefore, *to avoid path dependence, its possibility must be maintained*. In contrast, minimizing *potential* path dependence, creates sensitive to the initial game and inefficient outcomes.

Endogenous Institutional Change

Greif and Laitin (2004) describe a process of endogenous institutional change where game play creates a feedback that changes the payoff structure, what they call a *quasi-parameter*. To translate their quasi-parameter to our model, one can consider incremental adjustments to θ in each epoch. As θ changes, institutions can become reinforced or more fragile depending on whether the endogenous change is reinforcing or degrading - the latter occurs if the direction of the change in θ is toward the outcome that was initially inefficient. In our model, this means that the initial institution had an efficient outcome A but that as θ increases and crosses the efficiency cutpoint, B becomes efficient. In our model, behavior will only change once the quasi-paremeter enters the immune region. This implies inefficient outcomes for any games lying between the efficiency cutpoint and the immune region.

Claim 9. *A degrading quasi-parameter will produce behavioral change only when the quasi-parameter enters the immune region for the alternative strategy.*

The proof of the claim follows directly. Assume that all games produce outcome A. As θ increases, all outcomes will remain A until θ enters B's immune region. In other words, A is played in the entire susceptible region, beyond the efficiency cutpoint of $\theta^{=}$. A degrading quasi-parameter exemplifies *one-directional incrementalism*: a single behavior is reinforced with each change in θ.

Keep in mind that in Grief and Laitin's model, the quasi-parameter changes endogenously. We assume a sequence of exogenously chosen institutions. That distinction matters in so far as people's behavior under a new and similar institution differs from their behavior under an existing institution whose payoffs change. Specifically, one needs to make an assumption of the extent of cultural sway. If anything, one would expect that cultural sway, i.e. behavioral stickiness, would be larger for endogenous changes in payoffs to an existing institution than for a new, and similar institution. Thus, behavioral stickiness implies that endogenous degradation produces substantial inefficiency. In fact, it produces near maximal inefficiency as the equilibrium only changes once the quasi-parameter drifts into the immune region. To rectify this, our model would imply that one should speed up the degradation through a large change in quasi-parameter. This would move the game into the immune region for the efficient behavior.

4 Results for General Classes of Games

We now extend our model to cover all two by two symmetric games as well as a more general characterization for all games and strategies. We show that when cultural sway is large, new institutions with weak punishments for deviation are more likely to produce efficient outcomes. Creating strong rewards for choosing the correct strategy also improves the likelihood of efficient outcomes but is not as important as weak punishment for failure to coordinate.

Two by Two Symmetric Games

All two by two symmetric games can be parametrized as follows:

	A	B
A	ω, ω	ρ, ν
B	ν, ρ	$0, 0$

where the parameters ω, ν, and ρ can take any real value. There exist three possible efficient equilibria: A and B as before, and a third in which players alternate or switch between A and B, which we denote by S. The efficient regions for each of the three strategies are shown in Figure 4.

Our previous results require two modifications. First, we need a distance or similarity measure between games. One possibility would be a Euclidean distance measure on payoffs. Another approach would be to construct a lexicographic measure in which a game is closer to games that have the same efficient equilibrium behavior and then base distance on the Euclidean metric. The results that we derive do not depend on the distance metric used; we only need the measure to be well defined.

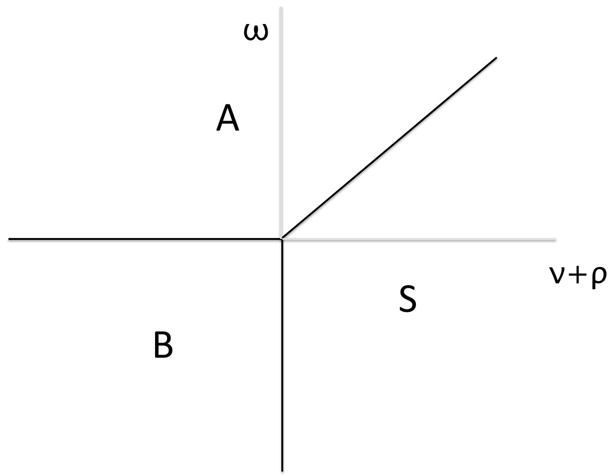

Figure 4: Efficient Regions For The Three Strategies in 2 x 2 Games

Second, we must now define punishment strategies. Consider the Prisoner's Dilemma game where A is the analog of cooperation. To support cooperation (A), an individual must punish with B in subsequent rounds of the game. We assume punishment relies on the minmax strategy.

Assumption 4 *Players choose strategies to achieve minmax payoffs when their opponent plays a different strategy.*

To see how path dependence arises in this setting, suppose that B is the efficient outcome in the first game and that the second game has the following payoffs:

	A	B
A	$2, 2$	ρ, ν
B	ν, ρ	$0, 0$

Assume $\rho + \nu < 4$, so that A is the efficient outcome. Let M denote the minmax payoff. Assuming the game is played repeatedly, we can approximate the payoffs from playing A, denoted as π_A, and the payoffs from playing strategy B, denoted as π_B, as follows:

$$\pi_A = \gamma M + (1-\gamma)\omega$$

$$\pi_B = \gamma 0 + (1-\gamma)M$$

The efficient outcome will arise in the second game if and only if $(1-\gamma)\omega > (1-2\gamma)M$. This will be satisfied if M equals zero or if $\gamma < \frac{1}{2}$ (given that $M \leq 0$). But suppose that the off-diagonal payoffs sum to a large negative number so that $M < 0$. Now, A may no longer be the equilibrium outcome in the second game. Thus, lowering the minmax payoff decreases the probability of getting the efficient strategy for the new game. *Stronger punishment is counterproductive.*

This intuition holds more generally. Suppose that we have an arbitrary family of games $G = \{G_\psi\}_{\psi \in \Psi}$ with a well-defined distance measure, $d : G \times G \to [0, \infty)$. Consider the introduction of game G_T in the Tth epoch. We can state the following claim:

Claim 10. *(***Stronger Punishment Impedes Efficiency***) Let G_τ denote the previous game in the sequence of games closest to G_T given d. Denote the payoff in the efficient repeated game equilibrium in G_T by A_T, A_τ denote the payoff in G_T from playing the equilibrium strategy used in G_τ, and let M denote the minmax payoff in G_T. The efficient equilibrium will be chosen in G_T if and only if the following holds:*

$$A_T > A_\tau + (A_\tau - M)\frac{(2\gamma - 1)}{(1-\gamma)}$$

The claim implies three routes to efficient outcomes: (1) make sure that the nearest game has the same efficient equilibrium, (2) increase the payoffs to the efficient equilibrium, or (3) increase the minmax payoff. This third insight is the most subtle. If a new institution creates large punishments, a small minmax payoff, then the cost of enticing those affected by cultural sway will be too high. Mild punishments, in contrast, enable the new efficient behavior to take hold. The idea that punishment works against innovation can also be found in Bednar (2009).

5 Discussion

Whether implementing a new law or managing a transition, be it on a grand scale, such as transitions to democracy, to more targeted goals like reducing obesity, the order that institutions are introduced can affect each one's performance. Scholars of development have noted the importance of institutional sequencing and the general sensitivity of institutional performance to the broader institutional context. Conflicting interpretations of the empirical evidence demands a new set of models of institutionalism to understand when and how institutional context and culture matters.

The growing interest in culture has lead to important empirical studies establishing a correlation between culture and institutional performance (Greif 1994, Guiso, Sapienza, and Zingales 2006, Tabellini 2010, Gorodnichenko and Roland 2013, Alesina and Giuliano 2013). In these studies, culturally-circumscribed attitudes are measurable proxies for equilibrium beliefs. Most scholars have zeroed in on the trust/distrust or individualistic/collectivist differences, as measured by the World Values Survey (Inglehart 1977, 1990, 1997). Analytically, culture has been treated as a primitive, at best "slow-moving" (Roland 2004). However, it is also the product of institutions (Putnam 1993, Tabellini 2010).

Our approach focuses on one component of culture—patterns of behavior—that may be more quickly subject to transformation than other aspects of culture, but have the same enduring qualities that could lead to suboptimal responses. For example, Becker and Woessmann 2009's analysis of Protestant success suggests that literacy—a behavior—and not (exclusively) a work ethic (a value or attitude) built the success of Protestant communities. "The linguistic and methodical skills created by the teaching of God's Word ... are thus valuable in other tasks beyond the religious realm" (Becker and Woessmann 2009:542). Beliefs are expectations of the behavior of others. Whether it is the practice itself or the expectation of the practice of others that is carried over is a subject for deeper study. In either case, institutions affect their development.

In the paper, we've first shown that if some individuals choose initial strategies based on their past experiences then we should expect to see path dependence in the performance of a sequence of institutions. In our model, path dependence arises under even mild spillovers. Relatedly, the level of inefficiency caused by this path dependence should correlate with the extent of cultural sway. Neither of these results should be especially surprising. Were the model not to produce such results, we would have reason to question the core assumptions.

The findings become less intuitive as we begin to explore comparative statics. As cultural sway increases, the susceptible regions increase and at some point, path dependence gives way to initial game dependence. If cultural sway is substantial, and if nearly all institutions have plausible behavioral analogues in the cultural repertoire, then the ultimate threshold will favor the behavior produced by the initial institution with high probability. The first institution, therefore, has an enormous effect. In light of this, potential path dependence implies contingency: multiple behaviors must be in play. The potential for multiple behaviors increase the likelihood of choosing the best action in future games. In contrast, a lack of path dependence along with a lack of behavioral diversity suggests lock-in and less efficiency.

These latter, less intuitive results, in their emphasis on maintaining the possibility of diverse behaviors, resonate with our results on the optimal sequencing of institutions. To maximize the likelihood of optimal responses to each institution, a sequence should start from diverse extremes (clear incentives to generate distinct behaviors) and gradually introduce institutions where outcomes are more contingent on the past. Thus, paradoxically, the way to reduce *realized* path dependence is to keep its *potential* alive for as long as possible. And the way to do that is to introduce diverse behaviors as early as possible. This will not occur through drift, or incremental, gradual change. Drift moves in a single direction and will only produce behavioral change when the institution moves into what we call the insulated region. Our model suggests that you cannot expect behavior to right itself until change is inevitable. Relatedly, when considering a more general class of games, we are able to show that weakening the ability to punish increases the likelihood of efficient outcomes. Weakening punishment encourages exploration by lowering its cost.

Our analysis produces two novel insights: First, if individuals learn from their past behavior, maintaining a diversity of strategies increases the likelihood of efficient outcomes by creating a system with greater adaptive capacity. Second, dependence on the path implies the presence of behavioral diversity capable of producing good outcomes. Thus, paradoxically, robustness—in this context we mean the ability to produce good outcomes in a variety of institutional settings—may require the potential for contingency.

Admittedly, models of microprocesses used to explain macrophenomenon will inevitably fail to capture important aspects of the environment that affect outcomes and limit our ability to construct effective institutions. The logic presented here suggests that when constructing institutions, one must take into account the behavioral repertoires of the relevant actors, and that doing so might help us to choose more effective sequencing of institutions.

References

Alesina, Alberto and Paola Giuliano. 2013. "Culture and Institutions." NBER Working Paper Series, Working Paper 19750.

Aoki, Masahiko. 1994. "The Contingent Governance of Teams: Analysis of Institutional Complementarity." *International Economic Review* 35(3):657–676.

Aoki, Masahiko. 2001. *Toward a comparative institutional analysis.* Cambridge, MA: MIT press.

Aoki, Masahiko. 2007. "Endogenizing institutions and institutional changes." *Journal of Institutional Economics* 3(1):1–31.

Axelrod, Robert. 1997. "The Dissemination of Culture: A Model with Local Convergence and Global Polarization." *Journal of Conflict Resolution* 41:203-226.

Becker, Sascha O. and Ludger Woessmann. 2009. "Was Weber Wrong? A Human Capital Theory of Protestant Economic History." *Quarterly Journal of Economics* 124(2):531–96.

Bednar, Jenna. 2009. *The Robust Federation: Principles of Design.* New York: Cambridge University Press.

Bednar, Jenna, Yan Chen, Xiao Liu, and Scott E. Page. 2012. "Behavioral spillovers and cognitive load in multiple games: An experimental study." *Games and Economic Behavior* 74(1):12–31.

Bednar, Jenna. and Scott E. Page. 2007. "Can Game(s) Theory Explain Culture? The Emergence of Cultural Behavior Within Multiple Games." *Rationality and Society* 19(1):65–97.

Berman, Sheri. 2007. "The Vain Hope for 'Correct' Timing." *Journal of Democracy* 18(3):14–17.

Boyd, Robert and Pete Richerson. 2005. *The Origin and Evolution of Cultures.* Oxford, UK: Oxford University Press.

Brady, Henry E. and David Collier. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards.* Lanham, MD: Rowman & Littlefield.

Camerer, Colin. 2003. *Behavioral Game Theory: Experiments on Strategic Interaction.* Princeton University Press, Princeton.

Carothers, Thomas. 2007. "The 'Sequencing' Fallacy." *Journal of Democracy* 18(1):12–27.

Cason, Tim, Anya Savikhin, and Roman Sheremeta. 2012. "Behavioral Spillovers in Coordination Games." *European Economic Review* 56:233–245.

Chwe, Michael S-Y. 2003. *Rational Ritual: Culture, Coordination, and Common Knowledge.* Princeton, NJ: Princeton University Press.

Dewatripont, Mathias and Gérard Roland. 1992. "The Virtues of Gradualism and Legitimacy in the Transition to a Market Economy." *The Economic Journal* 102(411):291–300.

Easterly, William and William Russell Easterly. 2001. *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics.* Cambridge, MA: MIT Press.

Easterly, William and William Russell Easterly. 2006. *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good.* New York: Penguin.

Falleti, Tulia G. and Julia F. Lynch. 2009. "Context and Causal Mechanisms in Political Science." *Comparative Political Studies* 42(9):1143–66.

Gilboa, Itzhak and David Schmeidler. 1995. "Case-based decision theory." *The Quarterly Journal of Economics* 110(3):605–639.

Gorodnichenko, Yuriy and Gerard Roland. 2013. "Culture, Institutions, and Democratization." University of California Berkeley manuscript. Available at `http://eml.berkeley.edu/~groland/pubs/gorrolpolculture03-05-2013.pdf.`

Greif, Avner. 1994. "Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies." *Journal of Political Economy* 102(5): 912-950.

Greif, Avner. 2006. *Institutions and the Path to the Modern Economy.* Cambridge, UK: Cambridge University Press.

Greif, Avner and David D. Laitin. 2004. "A Theory of Endogenous Institutional Change." *American Political Science Review* 98(4):633–652.

Goertz, Gary. 1994. *Contexts in International Politics.* Cambridge, UK: Cambridge University Press.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2006. "Does Culture Affect Economic Outcomes?" *Journal of Economic Perspectives* 20(2):23–48.

Henrich Joseph, Robert Boyd, Samuel Bowles, et al. 2001. "In search of homo economicus: Behavioral experiments in 15 small-scale societies." *The American Economic Review, Papers and Proceedings of the Hundred Thirteenth Annual Meeting of the American Economic* : 19(2):73–78.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Herbert Gintis, Ernst Fehr, and Colin Camerer, eds. 2004. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence in Fifteen Small-Scale Societies.* Oxford, UK: Oxford University Press.

Huntington, Samuel P. 1968. *Political Order in Changing Societies.* New Haven, CT: Yale University Press.

Hurwicz, Leonid. 1994. "Economic Design, Adjustment Processes, Mechanisms, and Institutions." *Economic Design* 1(1):1–14.

Inglehart, Ronald. 1990. *Culture Shift in Advanced Industrial Society.* Princeton, NJ: Princeton University Press.

Inglehart, Ronald. 1997. *Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies.* Princeton, NJ: Princeton University Press.

Inglehart, Ronald. (forthcoming) *Cultural Evolution.* manuscript.

Jehiel, Philippe. 2005. "Analogy-based Expectation Equilibrium." *Journal of Economic Theory* 123(2):81–104.

Johnson, James. 2014. "Models Among the Political Theorists." *American Journal of Political Science* forthcoming.

Knight, Jack and James Johnson. 2011. *The Priority of Democracy: The Political Consequences of Pragmatism.* Princeton, NJ: Princeton University Press.

Kranton, Rachel E. 1996. "Reciprocal Exchange: A Self-Sustaining System," *American Economic Review,*86(4):830–851.

Lake, David. 2010. "The Practice and Theory of US Statebuilding." *Journal of Intervention and Statebuilding* 4(3):257–284.

Licht, Amir N., Chanan Goldschmidt, and Shalom H. Schwartz. 2007. "Culture Rules: The Foundations of the Rule of Law and Other Norms of Governance." *Journal of Comparative Economics* 35:659–688.

Linz, Juan J. and Alfred Stepan, (1992) "Political Identities and Electoral Sequences: Spain, the Soviet Union, and Yugoslavia." *Daedalus* 121:123–139.

Linz, Juan J. and Alfred Stepan, (1996) *Problems of Democratic Transition and Consolidation: Southern Europe, South America, and Post-Communist Europe* Johns Hopkins University Press. Baltimore, MD.

Lipset, Seymour Martin. 1959. "Some Social Requisites of Democracy: Economic Development and Political Legitimacy." *American Political Science Review* 53(1):69–105.

Lipton, David and Jeffrey Sachs. 1990. "Creating a Market Economy in Eastern Europe: The Case of Poland." *Brookings Papers on Economic Activity* 1990(1):75–133.

Long, Norton E. 1958. "The Local Community as an Ecology of Games." *American Journal of Sociology* 64(3):251–61.

Mahoney, James. 2001. "Beyond correlational analysis: Recent innovations in theory and method." *Sociological Forum* 16(3):575–593.

Mahoney, James. 2010. "After KKV: The New Methodology of Qualitative Research." *World Politics* 62(1):120–47.

Mahoney, James and Kathleen Thelen. 2010. "A Theory of Gradual Institutional Change," in Mahoney J, Thelen K., eds., *Explaining Institutional Change.* New York: Cambridge University Press.

Mansfield, Edward D. and Jack Snyder. 2005. *Electing to Fight: Why Emerging Democracies Go to War.* Cambridge, MA: MIT Press.

Nash, John. 1951. "Non-cooperative games." *Annals of Mathematics* 54: 286–295.

Nisbett, Richard. 2003. *The Geography of Thought : How Asians and Westerners Think Differently . . . and Why.* New York, NY: Free Press.

North, Douglass C. 1993. "Institutions and credible commitment." *Journal of institutional and Theoretical Economics* 149(1):11–23.

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

Oyserman, Daphna and Spike W.S. Lee. 2008. "Does culture influence what and how we think? Effects of priming individualism and collectivism." *Psychological Bulletin* 134(2):311-342.

Page, Scott E. 2006. "Path Dependence." *Quarterly Journal of Political Science* 1: 87–115.

Putnam, Robert D. 1988. "Diplomacy and domestic politics: the logic of two-level games." *International Organization* 42(3):427–460.

Putnam, Robert. 1993. *Making Democracy Work: Civic Traditions in Modern Italy.* With Robert Leonardi and Raffaella Y. Nanetti. Princeton, NJ: Princeton University Press.

Roland, Gérard. 2000. *Transition and Economics: Politics, Markets, and Firms.* Cambridge, MA: MIT Press.

Roland, Gérard. 2002. "The Political Economy of Transition." *Journal of Economic Perspectives* 16(1):29–50.

Roland, Gérard. 2004. "Understanding institutional change: fast-moving and slow-moving institutions." *Studies in Comparative International Development* 38(4):109–131.

Samuelson, Larry. 1997. *Evolutionary Games and Equilibrium Selection* Cambridge, MA: MIT Press.

Samuelson, Larry. 2001. "Analogies, Adaptations, and Anomalies." *Journal of Economic Theory* 97:320-367.

Sen, Amartya. *Development as Freedom.* Oxford: Oxford University Press. 1999.

Swidler, Ann. 1986. "Culture in Action: Symbols and Strategies." *American Sociological Review* 51(2):273-286.

Tabellini. Guido. 2010. "Culture and Institutions: Economic Development in the Regions of Europe." *Journal of the European Economic Association* 8(4):677–716.

Taagepera, Rein and Matthew S. Shugart (1989) *Seats and votes: the effects and determinants of electoral systems.* Yale University Press, New Haven, Connecticut.

Talhelm, T., X. Zhang, S. Oishi, C. Shimin, D. Duan, X. Lan, and S. Kitayama. 2014. "Large-Scale Psychological Differences Within China Explained by Rice Versus Wheat Agriculture." *Science* 344(6184):603–608.

Thelen, Kathleen. 1999. "Historical Institutionalism in Comparative Politics." *Annual Review of Political Science* 2:369–404.

Tsebelis, George. 1990. *Nested Games: Rational Choice in Comparative Politics.* Berkeley, CA: University of California Press.

Tsebelis, George. 2002. *Veto Players: How Political Institutions Work.* Princeton, NJ: Princeton University Press.

Vermeule, Adrian. 2011. *The System of the Constitution.* Oxford, UK: Oxford University Press.

Weingast, Barry R. 1998. "Political Stability and Civil War: Institutions, Commitment, and American Democracy," in Robert H. Bates, et al, *Analytic Narratives.* Princeton, NJ: Princeton University Press.

Appendix

PF: Claim 1: Let $\pi_i(\theta)$ denote the payoff if both individuals choose strategy i and let π_{iD} denote the payoff to an individual who plays strategy i when the other player chooses the opposite. A game is immune for A if the payoff from A exceeds the payoff from B. If there exists no immune region, the result follows immediately. Therefore, assume an immune region for strategy A. The boundary of the immune region $\theta^A(\gamma)$ satisfies the following equation:

$$(1-\gamma)\pi_A(\theta^A(\gamma)) + \gamma\pi_{AD}(\theta^A) = (1-\gamma)\pi_{BD}(\theta^A) + \gamma\pi_B(\theta^A(\gamma))$$

Which can be rewritten as

$$\pi_A(\theta^A(\gamma)) - \pi_{BD}(\theta^A) = \frac{\gamma}{(1-\gamma)}\left[\pi_B(\theta^A(\gamma)) - \pi_{AD}(\theta^A)\right]$$

By construction, both individuals choosing strategy A is an equilibrium at $\theta^A(\gamma)$. Therefore, $\pi_A(\theta^A(\gamma) > \pi_{BD}(\theta^A)$. The fact that left hand side strictly exceeds zero implies that $\pi_B(\theta^A(\gamma)) > \pi_{AD}(\theta^A)$. If we increase γ to $\gamma + \epsilon$, This increases the coefficient on the right hand side of the equation. By A2, the left hand side of the equation increases as θ^A decreases and the right hand side decreases. Therefore, it follows that $\theta^A(\gamma+\epsilon) < \theta^A(\gamma)$. A similar argument holds for $\theta^B(\gamma)$ strictly increasing in γ.

PF: Lemma 1: It suffices to consider the case where $\theta_1 < \theta^=$. It follows that T_2 will equal θ^B as any susceptible game produces outcome A. Until there exists a k such that $\theta_k \geq \theta^B$, the threshold will remain at θ^B. Therefore, assume $\theta_2 < \theta^B$, then $T_3 = \theta^B$. If $\theta_2 \geq \theta^B$, then $T_3 = \frac{1}{2}(\theta_1 + \theta_2)$, provided that $\frac{1}{2}(\theta_1 + \theta_2)$ lies in the interval (θ^A, θ^B). If $\frac{1}{2}(\theta_1 + \theta_2) \leq \theta^A$, then $T_3 = \theta^A$, and if $\frac{1}{2}(\theta_1 + \theta_2) \geq \theta^B$, then $T_3 = \theta^B$. To determine the threshold for all subsequent periods, let θ^a equal largest θ_k for $k < t$ that produces outcome A let θ^b be the smallest θ_k for $k < t$ that produces outcome B. The threshold equals the average of θ^a and θ^b provided that lies in the susceptible region. Otherwise, it equals whichever of θ^A or θ^B is closest to that average.

PF: Claim 2: It suffices to show that the outcomes in the first two games in a sequence can differ. Let θ^S denote the susceptible game. Without loss of generality assume that A is payoff maximizing in the susceptible game. Let θ^o denote a game in which B is payoff maximizing. The sequence θ^o followed by θ^S produces outcome B in both games. The sequence θ^S followed by θ^o produces outcome A in the first game. The outcome in the second game will be A if $\theta^o < \theta^B$ and B otherwise.

PF: Corollary 1: Assume $\gamma < \hat{\gamma}$. It suffices to show that $T_t(\gamma) \leq T_t(\hat{\gamma})$ for all t. Let $\psi_t^A(\gamma)$ denote the largest θ_i for $i = 1$ to t that produces the outcome A given γ, and $\psi_t^B(\gamma)$ denote the smallest θ_i for $i = 1$ to t that produces the outcome B given γ. If there exists no θ_i that produces outcome B, set $\psi_t^B(\gamma) = \infty$. The proof relies on induction. By assumption $\gamma < \hat{\gamma}$.

Therefore by Claim 1, following period 1, three inequalities hold:

(i) $T_1(\gamma) < T_1(\hat{\gamma})$

(ii) $\psi_1^A(\gamma) \leq \psi_1^A(\hat{\gamma})$

(iii) $\psi_1^B(\gamma) \leq \psi_1^B(\hat{\gamma})$.

We will assume that all three inequalities hold through time t and show that they then hold for time $t+1$. There are three cases to consider.

Case 1: $\theta_{t+1} < \psi_t^A(\gamma)$ *or* $\theta_{t+1} > \psi_t^B(\hat{\gamma})$: By construction, $T_{t+1}(\gamma) = T_t(\gamma)$ and $T_{t+1}(\hat{\gamma}) = T_t(\hat{\gamma})$, so inequality (i) holds. Inequalities (ii) and (iii) hold because $\psi_{t+1}^j(\gamma) = \psi_t^j(\gamma)$ and $\psi_{t+1}^j(\hat{\gamma}) = \psi_t^j(\hat{\gamma})$ for $j = A, B$.

Case 2: $\psi_t^A(\gamma) < \theta_{t+1} < T_t(\gamma)$: First, consider the case where $\psi_t^B(\hat{\gamma}) = \infty$. In this case, $T_{t+1}(\hat{\gamma}) = T_t(\hat{\gamma}) = \theta^B(\hat{\gamma})$ (Recall that $\theta^B(\hat{\gamma})$ denotes the boundary for the immune region for B.) Therefore, by construction, $T_{t+1}(\gamma) \leq \theta^B(\gamma) \leq T_{t+1}(\hat{\gamma})$. The other two inequalities hold trivially. We can therefore restrict attention to the case where $\psi_t^B(\hat{\gamma}) < \infty$. By the induction hypothesis, $T_t(\gamma) \leq T_t(\hat{\gamma})$, therefore the outcome in the game θ_t is A for both spillover rates. Therefore, $\psi_t^B(\gamma) = \psi_{t+1}^B(\gamma)$ and $\psi_t^B(\hat{\gamma}) = \psi_{t+1}^B(\hat{\gamma})$ so (iii) holds. To see that inequality (ii) holds, note first that by assumption $\psi_{t+1}^A(\gamma) = \theta_{t+1}$. There are two possibilities to consider. First, if $\theta_{t+1} < \psi_t^A\hat{\gamma}$, then (ii) holds strictly. Otherwise, $\theta_{t+1} = \psi_t^A(\hat{\gamma})$ and $\psi_t^A(\gamma) = \psi_t^A(\hat{\gamma})$, so (ii) holds weakly.. To show that (i) holds, we first solve for $T_{t+1}(\gamma)$:

$$T_{t+1}(\gamma) = \frac{\theta_{t+1} + \psi_t^B(\gamma)}{2}$$

To solve for $T_{t+1}(\hat{\gamma})$, let $\theta^* = \max\{\theta_{t+1}, \psi_t^A(\hat{\gamma})\}$

$$T_{t+1}(\hat{\gamma}) = \frac{\theta^* + \psi_t^B(\hat{\gamma})}{2}$$

By the induction assumption, $\psi_t^B\gamma \geq \psi_t^B(\hat{\gamma})$ and by construction, $\theta^* \geq \theta_{t+1}$, which completes this case.

Case 3: $T_t(\gamma) < \theta_{t+1} < \psi_t^B(\hat{\gamma})$: First, consider the case where $\theta_{t+1} < T_t(\hat{\gamma})$, then the outcome is B for γ and A for $\hat{\gamma}$ and all three inequalities hold via straightforward arguments. If $\theta_{t+1} \geq T_t(\hat{\gamma})$, then the outcome is B for both γ and $\hat{\gamma}$. By assumption $\psi_{t+1}^B = \theta_{t+1}$ and

$$T_{t+1}(\hat{\gamma}) = \frac{\psi_t^A(\hat{\gamma}) + \theta_{t+1}}{2}$$

To solve for $T_{t+1}(\gamma)$, let $\theta^* = \min\{\theta_{t+1}, \psi_t^B(\gamma)\}$

$$T_{t+1}(\gamma) = \frac{\psi_t^A(\gamma) + \theta^*}{2}$$

By induction, $\psi_t^A\gamma \leq \psi_t^A(\hat{\gamma})$ and by construction, $\theta^* \leq \theta_{t+1}$, which completes the proof.

PF: Claim 3: The proof is by counterexample using payoffs from the traditional and innovative strategies game. Assume context $\Omega = \{0.8, (0.1, 0.9)\}$ and context $\hat{\Omega} = \{0.75, ()\}$. Initially, $T = \hat{T} = 8$. Given this construction, the susceptible region in context Ω will be larger than the susceptible region for context $\hat{\Omega}$. The continuation (0.1), will have no effect on the context Ω, but moves the threshold in context $\hat{\Omega}$, $\hat{T}$ to 10. Therefore, Ω cannot be more path dependent.

PF: Claim 4: By next outcome equivalence, the two contexts have the same threshold. Denote these by T and $\hat{T}$. Let θ^a equal the largest θ_k in context Ω that produces outcome A. Define $\hat{\theta}^a$ similarly for context $\hat{\Omega}$. Similarly, let θ^b equal the smallest θ_k in context Ω that produces outcome B. The interval $[\theta_L, \theta_U]$ can be partitioned into six intervals $[\theta_L, \theta^a)$, $[\theta^a, \hat{\theta}^a)$, $[\hat{\theta}^a, T)$, $[T, \hat{\theta}^b)$, $[\hat{\theta}^b, \theta^b)$, and $[\theta^b), \theta_U]$.

Without loss of generality, assume that for the next game that arises $\theta < T$. The outcome in that game will be A. We first state a lemma that simplifies the remainder of the proof.

Lemma 2. *The introduction of the first new game moves T, the threshold in context Ω, at least as far as it moves $\hat{T}$, the threshold in context $\hat{\Omega}$.*

The proof considers distinct cases. Note that if context Ω has produced a B outcome, then so has $\hat{\Omega}$. If $\theta \in [\theta_L, \theta^a)$, then neither threshold moves so the result holds. If $\theta \in [\theta^a, \hat{\theta}^a)$, then only T moves, so the result holds. Finally, if $\theta \in [\hat{\theta}^a, T)$, then the thresholds move to $\frac{\theta+\theta^b}{2}$ and $\frac{\theta+\hat{\theta}^b}{2}$ in contexts Ω and $\hat{\Omega}$ respectively. Given that $\theta^b \leq \hat{\theta}^b$, the result follows.

Given the lemma, it follows that after the introduction of the game θ, the set of games that will now have different outcomes is larger in context Ω that in context $\hat{\Omega}$. This follows because the threshold has moved further in context Ω. Therefore, after one game has been added, context $\Omega produces more path dependence than$ context $\hat{\Omega}$. Note that Following any continuation, the susceptible region of Ω is at least as large as the susceptible region of $\hat{\Omega}$. We now state another lemma:

Lemma 3. *If contexts Ω and $\hat{\Omega}$ have both produced both types of outcomes and if Ω has a larger susceptible region, then any new game will move T at least as far as it moves $\hat{T}$*

pf. Assume that in the new game $\theta < T$. Suppose first that $T \leq \hat{T}$. If $\theta \in [\theta_L,)$, then $\hat{T}$ does not change, so the result holds. If the interval $[\hat{\theta}^a, T)$ is not empty and contains θ then the thresholds become $\frac{\theta+\theta^b}{2}$ and $\frac{\theta+\hat{\theta}^b}{2}$ in contexts Ω and $\hat{\Omega}$ respectively. Given that $\theta^b \leq \hat{\theta}^b$, the result follows.

Next suppose that $T \geq \hat{T}$. As before, if If $\theta \in [\theta_L, \hat{\theta}^a)$, then $\hat{T}$ does not change as before, so the result again holds. If $\theta \in [\hat{\theta}^a, \hat{T})$ then the thresholds move to $\frac{\theta+\theta^b}{2}$ and $\frac{\theta+\hat{\theta}^b}{2}$ in contexts Ω and $\hat{\Omega}$ respectively. Given that $\theta^b \leq \hat{\theta}^b$, the result follows. Finally, suppose that $\theta \in [\hat{T}, T)$. Now the outcomes in the two contexts differ. The outcome in context Ω is A but the outcome in context $\hat{\Omega}$ is B. The thresholds therefore move to $\frac{\theta+\theta^b}{2}$ and $\frac{\theta+\hat{\theta}^a}{2}$ in contexts Ω and $\hat{\Omega}$ respectively. In context Ω, the threshold moves a distance $\frac{1}{2}(\theta - \theta^a)$. In context $\hat{\Omega}$, the threshold moves a distance $\frac{1}{2}(\theta - \hat{\theta}^b)$. Given that $\hat{T}$ is the midpoint of $\hat{\theta}^a$ and $\hat{\theta}^b$, the result follows from the fact that $\mid \theta - \hat{\theta}^b \mid < \mid \theta - \hat{\theta}^a \mid$ and that $\theta^a < \hat{\theta}^a$.

PF: Claim 5: By Claim 1, the size of the initial susceptible region weakly increases in γ. To show that initial path dependence strictly increases, it suffices to show first, that for any continuation $(\theta_1, \theta_2, ...\theta_k)$, that if all outcomes are the same given γ, then they must also all be the same for $\hat{\gamma} > \gamma$, and second, that there exists a continuation that produces a different outcome given γ but not given $\hat{\gamma}$. It suffices to show for the case where the first outcome is A. Note that in any continuation all outcomes are A if and only if $\theta_i < \theta^B(\gamma)$, the boundary of the immune region for B given γ. The result follows from the fact that that $\theta^B(\hat{\gamma}) > \theta^B(\gamma)$. To show that there exists a continuation that produces an outcome of B for some game under γ but not under $\hat{\gamma}$, consider the single game continuation, $\theta_2 \in (\theta^B(\hat{\gamma}), \theta^B(\gamma))$. It has outcome B in the context defined by γ and outcome A in the context defined by $\hat{\gamma}$.

The proof that in the limit as γ approaches one, that the extent of initial game dependence converges to one, follows directly from [A1] and [A2].

PF: Claim 6: To simplify notation, we write $\theta^B(\gamma)$ as θ^B and define θ^A similarly. Choose θ_1 in the interval $(\theta^A, \theta^=)$ and θ_2 in the interval $(\theta^=, \theta^B)$. By construction, both are susceptible. In the sequencing (θ_1, θ_2), A will be chosen in both games, resulting in an inefficient outcome in game θ_2. In the sequencing (θ_2, θ_1), B will be chosen in both games, resulting in an inefficient outcome in game θ_1.

PF: Claim 7: We first prove that the condition is sufficient. Suppose no games exceed balanced sequencing. It suffices to consider the case where $R < M$. Then the sequence, $(\alpha_1, \beta_1, \alpha_2, \beta_2, ...\alpha_R, \beta_R, ...\beta_M)$ results in efficient outcomes for each game. In what follows, we refer to this as the *alternating sequence.* When game α_j occurs in the sequence $j-1$ of the β games have been added to the sequence. By assumption $j - 1 < I(\alpha_j)$, which implies that the efficient outcome occurs in game α_j. Similarly, when β_i occurs in the sequence i of the α games have been added to the sequence. By assumption $i \leq I(\beta_i)$, which implies that the efficient outcome occurs in game β_i.

Next assume that there exists games that exceed balanced sequencing. Let $I(\alpha_{j'})$ have the smallest value among the α's that exceed balanced sequencing and let $I(\beta_{i'})$ have the smallest value among β's that do. Note first that $I(\alpha_{j'})$ cannot equal $I(\beta_{i'})$. If it did, given that $i' > I(\beta_{i'}) = I(\alpha_{j'})$, which by condition (1) implies that $I(\beta_{i'}) \geq j'$, but $I(\beta_{i'}) = I(\alpha_{j'})$ which by assumption is strictly less than j', a contradiction.

The algorithm for adding games consists of two steps. We write it assuming that $I(\alpha_{j'}) < I(\beta_{i'})$. The other case is symmetric.

Step 1: Introduce all games up to game $I(\alpha_{j'})$ in both sequences according to the alternating sequence. For example, if $I(\alpha_5) = 3$, then introduce games $\alpha_1, \beta_1, \alpha_2, \beta_2, \alpha_3$, and β_3. These games will all produce efficient outcomes.

Step 2: Add the α games as they appear in the alternating sequence through $\alpha_{j'}$.

Step 3: If none of the remaining games exceed balanced sequencing add them in the order in which they appear in the alternating sequence. If not, choose the unique game with the smallest index that exceeds balanced sequencing and go to Step 1.

To see that this produces efficient outcomes, note first that for the games up to game j' in both sequences, their indices exceed their subscript so the efficient outcomes occur. In Step 2, the efficient outcomes also occur in games α_j through $\alpha_{j'}$. In Step 3, suppose first that no other games satisfy the conditions. By (1), if $i > I(\alpha_{j'})$, then $I(\beta_i) \geq j'$, so games β_i for $i = I(\alpha_{j'}$ to j' produce efficient outcomes. Suppose instead that there exists another game that satisfies one of the conditions. The games added in Step 1 might consist only of β's. If so the logic just described applies. If it also consists of some α's then those α's all have subscripts that exceed their indices implying that they produce efficient outcomes.

To prove necessity, suppose that the conditions are violated. Let $\hat{j}$ equal the smallest j that exceeds balanced sequencing. Define $\hat{i}$ similarly if it exists. Assume that $\hat{j} \leq \hat{i}$. The other case is symmetric. By our assumption that the conditions are violated, there exists a β_i s.t. $i > I(\alpha_j)$ with $I(\beta_i) < j$ Given a sequencing of the games, suppose that $\alpha_{\hat{j}}$ comes before β_i. By assumption, $I(\beta_i) < \hat{j}$, which implies that β_i produces the inefficient outcome. Alternatively, suppose that β_i occurs before α_j. By assumption $i > I(\alpha_j)$, which implies that the inefficient outcome occurs in game α_j which completes the proof.

PF: Claim 8: Assume a $j > i$ s.t. α_j appears before α_i, and that this is the last pair such that this condition holds. Let t denote that epoch in which game α_j appears.

Case 1: outcome in game α_j is A. The outcome in game α_i must also be A. In this case, move α_i in front of α_j in the sequence. Given that the outcome in game α_j equals A, $T_t > \alpha_j$. Given $\alpha_i < \alpha_j$, the outcome in game α_i will be A in the new sequence. Furthermore, T_{t+1} will be unchanged, so outcomes in all other games will be unchanged.

Case 2: outcome in game α_j is B and the outcome in game α_i is A. Move all games labeled as β that occur after α_j ahead of α_j keeping their relative order unchanged and then switch the order of α_j and α_i. By doing this, any games whose order has been changed prior to game α_i are labeled with β's and occur in period t or later. Therefore, all thresholds up through T_t are unchanged. Bame α_j in the original sequence produced outcome B, so in the original sequence, $T_t < \theta^=$. Therefore, the outcomes for all games that appear between epoch t and game α_i in the new sequence will be B which is efficient. Therefore, the number of efficient outcomes for games labelled with β that occur after epoch t cannot decrease in the new sequence.

Next, consider game α_i. If it lies in the immune region, then its outcome will be efficient in the new sequence. Suppose not. In the original sequence, it's outcome was A. And, at the time α_i occurred, the threshold was strictly less that α_j. (Follows from the fact that α_j produced an outcome of B in the original sequence.) Given α_j not immune, from lemma 1 we know that the threshold at the time that game α_i arrived in the original sequence equaled the average of the largest θ that produced outcome A and the smallest θ that produced outcome B. The set of games that produce outcome A that occur prior to game α_i remains unchanged, but the set of games that produce outcome B that appear prior to game α_i no longer contains game α_j and now includes all games labelled with β that occurred after α_i in the original sequence. Therefore, the game with the smallest θ that produces outcome B in the new sequence that occurs prior to α_i is weakly larger than the smallest θ that produced outcome B in the original sequence. It follows that, the threshold that exists prior to game α_i in the new sequence is higher than in the original sequence so the outcome in game α_i remains A.

Next consider the outcome for game α_j. It was previously B. Given that game α_j now follows game α_i, the outcome in game α_j could now become A, in which case the outcomes are more efficient. If not, outcomes are not less efficient. Finally, consider the games that follow game α_j in the new sequence. Assume first that the outcome in game α_j remains B. We will show that the threshold it faces is unchanged. To do so, we must first state a lemma.

Lemma 4. *There cannot exist a sequence of games such that a game labelled as an α produces an outcome B and a game labelled as a β produces an outcome A.*

pf. By symmetry we need only consider the case where the first game in the sequence to produce an inefficient outcome has a label α. Denote this game by α^*. From lemma 1, the threshold in all remaining games will be less than or equal to α^*. Therefore, a game labelled as a β that occurs later in the sequence has a $\theta > \theta^= > \alpha^*$ and produces outcome B. By assumption, any game labelled as a β that occurs earlier in the sequence produces an outcome of B, which completes the proof.

By the previous lemma, given that game α_j produces outcome B, only games labelled as α can produce inefficient outcomes in either sequence. Furthermore, at that point in the sequence where there exist previous games in the sequence labelled as α that produce both outcomes, the threshold will be the average of the largest α that produces an outcome of A and the smallest α that produces an outcome of B.

Consider the game that occurs immediately after α_j in the new sequence, call this game α_τ. The threshold at the time that this game arises in the original sequence of games equals the average of the largest α_k corresponding to an earlier game that produced outcome A and the largest $\alpha_{k'}$ corresponding to an earlier game that produced an outcome B. Note first that by assumption, the outcomes in games α_i and α_j are unchanged in the two sequences. Note next that all other games labelled as α that occur prior to α_τ have identical prior games in the two sequences, so their outcomes must also be unchanged. Therefore, the threshold in effect at the time that α_τ arises is the same for the two sequences. By induction, all remaining games also face the same thresholds.

Last, suppose that the outcome in game α_j becomes A in the new sequence. This is the efficient outcome. Consider the game, α_τ that follows the game α_j in the new sequence. If it's threshold is unchanged, then there exists a game $\alpha_m$that occurs prior to α_j in both sequences, that is labelled as an α, produces outcome B, and satisfies $\alpha_m \leq \alpha_j$. If such an α_m exists, then by the argument in the previous paragraph, outcomes in α_τ and all subsequent games are unchanged. Suppose that no such α_m exists. It follows that in the original sequence α_j was the smallest θ producing an outcome of B. This implies that the threshold for game $\alpha_t au$ will be higher in the new sequence than the original sequence. An inductive argument shows that the same result holds for all subsequent games in the new sequence. Therefore, at least as many games in the new sequence must produce the efficient outcome, thus completing the proof.

Proof of Claim 10: The payoff from playing the efficient strategy in G_T equals $\gamma M + (1-\gamma)A_T$. The payoff from playing the equilibrium strategy used in the G_τ equals $\gamma A_\tau + (1-\gamma)M$. The first expression is larger than the second if and only if $(2\gamma - 1)M + (1-\gamma)A_T > \gamma A_\tau$. This can be rewritten as $(2\gamma - 1)(A_\tau + M - A_\tau) + (1-\gamma)A_T > \gamma A_\tau$. Rearranging terms gives the result.