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

Jenna Bednar and Scott E Page*

September 20, 2015

Abstract

To understand how different rules, laws, and other institutions affect one another's performance, we propose a model of *behavioral spillovers* across institutions. Agents' initial reaction to an institution depends on their existing behavioral repertoires. We find that as spillovers increase in magnitude, path dependence becomes likely; however, when spillovers become large, the path becomes less important than the initial choice of institution. We use this model to characterize the optimal sequencing of institutions. We find that the best-performing sequences initially induce behavioral diversity and then build on existing productive behaviors to avoid inefficient spillovers. Counter-intuitively, these sequences maximize *potential* for path dependence in order to avoid its *realization*. We also characterize conditions that make institutions less prone to interaction effects and find that optimal payoff structures rely on weak punishments.

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; comments from David Laitin, Tim van Zandt, and Jenna's fall 2014 Institutional Analysis class; audience members at the Priorat Conference on Political Institutions, INSEAD, Hamburg University, University of Pennsylvania, the Université de Génève, ETH-Zurich, the University of Michigan, the Santa Fe Institute, and Princeton University as well as research support from the U.S. Army Research Laboratory and the U.S. Army Research Office under contract/grant number W911NF1010379.

Societies implement formal institutions—rules and laws—to shape behavior in political, social, and economic activities. Although institutions are in theory designed with behavioral goals in mind, in practice, institutions often produce outcomes that differ from aspirations, sometimes wildly so. Moreover, similar institutions operating in different contexts can perform differently. Institutional performance is a function of context.

To explain the gap between theory and experience, we hold constant many standard features of the environment—resources, hostile neighbors, diverse goals—to focus on the role played by preexisting institutions. It is intuitive to think that a person's behavior within an institution depends on her belief system, history, cultural background, and her expectations of how others will behave. It is equally intuitive that existing institutions play some role in constructing that context. But the devil of true understanding lies in the details. Intuition doesn't explain *how* context matters and thus can't explain *why* it influences outcomes. To understand how context affects outcomes, one must identify specific mechanisms. We can then use models to identify conditions for institutional interdependence to have meaningful effects.

Many studies focus on *direct* interactions, examining how multiple institutions co-constrain or jointly motivate a particular choice, as in 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). We focus instead on *indirect* interactions that occur as the behaviors induced by one institution spill over to affect agents' responses to a different institution.

We construct a general framework where the ensemble of institutions creates a behavioral context. This context influences the performance of subsequent institutions in a manner that we capture theoretically, as follows: Institutions induce behaviors, including simple cooperation strategies as well as more elaborate protocals for producing trust or exacting punishment.

These behaviors become part of the environment where other institutions operate, and are possible initial responses when new institutions are introduced. The predisposition to initially choose a known behavior can either interfere with or enable the successful operation of a new institution. Thus, how an institution performs can depend on the behaviors produced by the *ensemble* of existing institutions and the behaviors they induce.

We've organized this paper into five parts. We first situate our project within the literature on institutional design and provide a preview of our results. In Section 2, we present our modeling framework and apply it to two families of games that can be parameterized by a single variable. We also include a sketch of how one might apply the model to sequencing in democratic transitions. We present our main results in Section 3, parsing path dependence from initial game dependence, optimal sequencing, and endogenous institutional change. In the fourth part, we extend the model to cover a broad array of cooperation problems: in formal modeling terms, we present a model of 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 Motivation and Preview of Results

Many scholars concerned with institutional performance have paid attention to institutional context. 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 his approach to institutional change also allows for interdependence between institutions (Aoki 2007). North's (1993) "institutional matrix" interacts with beliefs to restrict institutional change to incremental advances. In Mahoney and Thelen (2010), the dynamics are also explicit and agency influences incremental institutional change.

Formal analysis of optimal sequencing requires situating an institution within an 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). 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, Acemoglu and Robinson 2012) and in studies of endogenous institutional change (Greif and Laitin 2004, Greif 2006, Mahoney and Thelen 2010).

Theories of democratization and economic development imply sequencing the introduction of institutional reform, and sequencing prescriptions conflict. Consider 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 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 reduction of economic inequality (Boix 2003), and an inclusive political economy (Acemoglu and Robinson 2012). Sometimes recommendations prioritize security and order above all other objectives (Mansfield and Snyder 2005, Lake 2010). Despite their theoretical diversity—and there are dozens of policy analyses that take a position on this topic—these works share two features: they are empirically grounded and they all find that institutional order affects outcomes. They just disagree about the optimal ordering.

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.

Behavioral spillovers are the core assumption of our model.[1] Individuals interact across multiple institutional settings, and the behaviors that emerge in any one context—be they cooperative, trusting, altruistic, or competitive—might bleed into other institutional settings, creating a consistency of behavior across contexts as well as path dependence.[2] Given that behavioral spillovers affect outcomes in other institutions, they provide a logic for how the extant institutions can either facilitate or undermine the success of new institutions.

[1]Support for the existence of behavioral spillovers is 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). Within political science, reliance on past experiences and habits can be found in Finnemore and Sikkink's (1998) explanation of internalization. At a more macro level, the assumption of spillovers producing consistency also aligns with cross-national survey research on cultural diversity (Inglehart 1990, 1997).

[2]Cross-institutional behavioral consistency is not a given. Agents observe what others have done and will copy a higher-earning behavior if one exists. Thus, our assumption of initial actions based on the past creates the possibility of consistency but in no way guarantees it.

Communities that can leverage existing behaviors optimally as they encounter new institutions will do well. Those that cannot may fare poorly. Outcomes within our framework will often be inefficient. The existence of spillovers also points to a shortcoming with the standard approach of considering institutions in isolation of context. Stated more strongly, in light of spillovers, one can read the institutional design and analysis literature as partial equilibrium analysis.

Our framework relies on four primary assumptions. First, individuals have *repertoires* of behaviors that they acquire in response to institutions. Second, these behavioral repertoires develop sequentially as individuals encounter more institutions. Third, how individuals behave initially in a novel game depends on the extent of *cultural sway.* Some agents will respond to a new institution with a behavior they use in similar situations.[3] Individuals not affected by cultural sway approach the new game with a blank slate, focusing entirely on the payoff structure. This leads them to choose the payoff maximizing equilibrium action initially. Fourth, after their initial responses to a new institution, individuals learn new behaviors by best responding to the behaviors of others.

While we focus on behavior, there exist alternative approaches. Historical narratives situate institutions' contextual effects in beliefs, behaviors, norms, rituals, habits, and organizations (Greif 2006), but any formal model must reduce the dimensionality of causes. Greif, for example, relies on *beliefs* as the cultural attribute that transmits the weight of past institutions and constrains the set of equilibria as well as determining public acceptance of institutions (Roland 2000). Although behavior depends on beliefs, no one-to-one mapping exists between the two. Common beliefs need not induce identical behaviors and behavioral heterogeneity can have implications for outcomes (Bednar et al 2015). Alternatively, identi-

[3] Behavioral consistency is an often-mentioned component of culture. Culture consists of a much larger bundle of socially-transmitted elements, including values, beliefs, traits, narratives, and the meaning of symbols and artifacts. See for example Swidler 1986, Axelrod 1997, Boyd and Richerson 2005 and Bednar and Page (2007). The observability of behavior relative to preferences, shared meanings, or beliefs, advantages its use in empirical settings.

cal behaviors can emerge despite disparate beliefs. While both beliefs and behavior can be used to identify conditions for institutional path dependence, they rely on different assumptions. Belief-based models require constraints on priors, while our model requires minimal bounds on the extent of the cultural sway. A behavioral approach complements belief-based models by providing an opportunity to explore a different set of causal forces and to draw distinct insights.

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" (2006:191). 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 to evaluate the gradualist approach. We can isolate the conditions when gradualism would and would not lead to optimization of social goals.[4] Gradualism generically leads to inefficient outcomes by locking in on a particular behavior. 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 the dimensionality of behavioral repertoires, resulting in greater capacity to respond optimally to incentive structures. Surprisingly, this initial diversity maximizes the potential for path dependence, but, importantly, that dependence is not realized. What will be realized are

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

efficient outcomes through incremental changes to behaviors drawn from a diverse repertoire.

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 that as cultural sway grows in influence, path dependence decreases, and initial institutional choice can determine outcomes. Finally, we show in a general class of games that optimal institutional design implies more carrots than sticks, with 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 optimal sequences enable the possibility of path dependence but then avoid it. Third, linking our model to the quasi-parameter model of Greif and Laitin (2004), we find that negatively reinforced institutional drift leads to institutional change at an inefficient moment.

2 The Model

Our framework relies on three assumptions: (1) institutions arrive sequentially, (2) individuals' initial behaviors differ: some draw on past behaviors and others play the payoff maximizing strategy; and (3) in subsequent periods, individuals learn to play equilibrium in the new institution.

We assume an infinite population of individuals who play a sequence of games. We divide our population into two categories: those whose initial action is *culturally embedded* in current practices and those behaviors are *context free*.[5] 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 reason by analogy.[6] 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. Note that the context-free response is an implicit assumption in many, if not most, formal models of institutions.

To clarify the presentation, 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 is 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 so, we assume 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 and the payoff maximizing repeated game equilibrium strategy by s_t^*.[7] In the first epoch, we assume that all individuals choose s_1^*, the 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).

We define the *cultural sway* to be 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), identify the game in the sequence that most closely resembles game g_t, and initially play that strategy in the new

[5]In experimental settings, initial game play is heterogenous (Camerer 2003).

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

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

game.[8] The remaining fraction $(1-\gamma)$ of the individuals choose context-free behavior, which we assume to be the payoff maximizing equilibrium strategy, s_t^*.

A central part of our analysis will be the extent to which a sequence of games together with a spillover parameter enable path dependence. To simplify the presentation, we define an **historical context** to be an initial history of games together with a spillover parameter: $\Omega = \{\gamma, (g_1, g_2, ..g_k)\}$. Without loss of generality, assume a game g that when played given a historical context produces an efficient outcome. Next, imagine inserting a sequence of games between the history of games and game g. The outcome in g exhibits path dependence (relative to the context) if there exist sequence insertions that can change the outcome in game g. In this case, that would mean making the outcome in g inefficient.

We can compare relative degrees of path dependence in the following way. Historical context Ω is **more path dependent** than $\hat{\Omega}$ if (1) both produce the same outcome in game g, and (2) the set of sequence insertions that change the outcome in context Ω strictly contains the set of sequence insertions that change the outcome in context $\hat{\Omega}$. Put another way, outcomes in the context Ω are less robust to the insertion of sequences than in context $\hat{\Omega}$.[9] That is, more inserted sequences would switch the outcome in g given Ω than given $\hat{\Omega}$.

In the next section, we show that the performance of some institutions (game forms) depends on the institutions that were introduced prior to its appearance. We refer to these institutions as *susceptible*: behavioral outcomes depend upon the particular sequence of games that precede it. If a game's outcome is not a function of the historical context, we refer to it as *immune*. As we will see, immunity is harder to achieve in contexts with substantial cultural sway. In addition, the initial game in the sequence can have a large effect on future outcomes. We define the *extent of initial game dependence* for a context Ω

[8]Identifying the nearest prior institution requires a distance function between games, which we define below.

[9]For formal definition see appendix.

to be the probability that the outcome of a game in the susceptible region is the same as that of the initial game in the context.

Tradition and Trust

To build an intuition for the primary concepts in our model, we derive results for two classes of games that can be parameterized along a single dimension. The first class of games applies to situations in which individuals must choose to act traditionally or to innovate. The second applies to situations in which people can take a trusting action that exposes them to a possible loss or take a safe action.

In the first class of games, individuals either *follow tradition* or *innovate*. 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. If the two players choose opposite actions then each receives a payoff of four. For θ less than four or greater than twelve, the game has a unique equilibrium. For $\theta \in [4, 12]$, both sticking to tradition (T) and innovating (I) are pure strategy equilibria. Note that sticking to tradition is efficient if $\theta \leq 8$ and innovating is efficient if $\theta \geq 8$.

Figure 1: Payoffs for the Tradition/Innovation Game

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

To demonstrate the logic of the model, we 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. First, 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 to follow tradition. Assume that in the second game $\theta_2 = 9$. By construction, three-fourths of the population

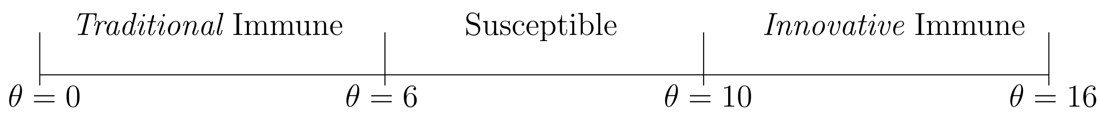

Figure 2: Susceptible and Immune Regions in the Tradition/Innovative Game as a Function of θ

will initially follow tradition and one-fourth will innovate. The payoffs for the two strategies in the population are as follows:

$$\textit{Tradition (T):}\ \tfrac{3}{4}(7) + \tfrac{1}{4}(4) \ = \tfrac{25}{4} \qquad \textit{Innovate (I):}\ \tfrac{3}{4}(4) + \tfrac{1}{4}(9) \ = \tfrac{21}{4}$$

For $\theta = 9$, if everyone were to innovate, they would earn higher payoffs, but the payoff from sticking to tradition 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 to follow tradition.

Alternatively, if the first game in the sequence had produced innovative strategies, i.e. $\theta_1 > 8$, the outcome in the second game with $\theta_2 = 9$ would also have been to innovate. 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 exhibit true path dependence: they depend not just on the set of games, i.e. *set dependence*, but also on the order in which those games are played (Page 2006).

Not all games will be susceptible. If θ is sufficiently high (resp. low) then the outcome will be to innovate (resp. follow tradition) regardless of the previous games, as depicted in Figure 2. 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.[10] A similar calculation shows that for $\theta_t < 6$, the strategy chosen will follow tradition regardless of the previous games played. Therefore, the values $\theta = 6$ and $\theta = 10$ partition the parameters into the *immune* and *susceptible* regions.

We next consider a family of trust games with a *Safe* action and a *Trusting* action. This family generalizes the Stag Hunt game. Given the payoffs, if $\theta \leq 8$ then Safe is the efficient equilibrium, otherwise Trusting is efficient.

Figure 3: Payoffs in the Trust Game

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

The initial susceptible regions are as shown in Figure 4.[11]

For this class of games, the immune regions favor the Safe action because it is *risk dominant*. Learning advantages risk dominant strategies (Samuelson 1997). Trust, therefore, will be harder to sustain. Consider the following set of games $\{7, 9, 10, 11, 14\}$. If game $\theta = 7$ occurs first, then the only sequence that obtains the efficient outcomes in all remaining games is $(7, 14, 11, 10, 9)$. This sequence frontloads games in which Trusting is the efficient outcome and then builds trust in the susceptible region.

[10]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 innovation 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 $\theta_2 \geq 10$.

[11]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 the threshold for the safe action as $\theta = 6.5$.

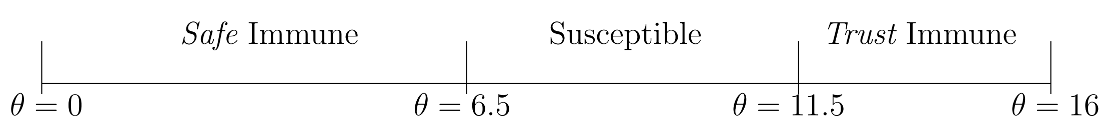

Figure 4: Susceptible and Immune Regions in Safe/Trusting Game as a Function of θ

N Player Games: Sequencing of Electoral Institutions

The classes of games considered so far assume two player games. Our framework also applies to games with arbitrary numbers of players. Consider an N person coordination model in which voters 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, voters would be more likely to coordinate on regional parties. When national elections are held, voters may continue to support regional parties. In contrast, if the national elections were held first, national parties may be more likely to emerge. Later, when subsequent regional elections are held, those nationalist actions would spill over into the regional election. Relevant behaviors in this game could include gathering information, developing policy platforms, and forming relationships with people outside the region. These behaviors might transfer to the other elections.[12]

Linz and Stepan (1992, 1996) make similar arguments (without relying on a formal model) to advocate that new democracies hold national elections first. Spain, where national parties won a majority of the vote in early elections despite strong Basque and Catalan regional identities provides a supporting example. In Yugoslavia and the former Soviet Union regional elections were held first and the results were less successful.

Our framework reveals the conditionality of their claims. If the payoffs from coordinating on regional interests in say Moldavia, Georgia, and Ukraine were sufficiently strong, region-

[12] A formal version might look as follows: assume N voters within a region who can 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 regional focus in the regional election and national focused in the national election.

alist behavior could have been 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 later. Linz and Stepan's argument that holding national elections first would have solved the problem assumes a limited attachment to regional identities. In specifying conditions, our model helps to build a testable hypothesis of their argument.

3 Results: One-Dimensional Family of Games

We now state general results for a family of games that includes coordination games and trust games. 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.*[13]

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. $\theta^B = \theta_U$).

[13]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.

Our first claim states that the size of the initial susceptible region increases in the size of the spillover: the stronger the spillover, the more likely inefficient equilibria emerge in later games. The proofs of 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.*

Next, we state a lemma that clarifies the logic. The lemma 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 $\theta > T$. Note that the lemma implies that two historical contexts are outcome equivalent if and only if they have the same threshold.

Lemma 1. *The outcome in a game is determined by a threshold in the space of payoffs that depends on the historical context and the amount of cultural sway. [Given a historical context Γ of length $t-1$, in epoch t there exists a threshold $T_t(\Gamma)$ such that if $\theta_t < T_t$, A will be the outcome and if $\theta_t > T_t$, B will be the outcome.]*

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 larger the effect of the initial game.

Corollary 1. *If the initial game produces outcome A, then for any subsequent games, the threshold increases in the amount of cultural sway and in the payoff parameter of the initial game. [Given $\Omega = \{\gamma, (\theta_1)\}$, where $\theta_1 < \theta^=$, for any sequence of future games $(\theta_2, \theta_3, \ldots, \theta_k)$, 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 institutional path dependence depends on historical context. We first state a sufficient condition for the existence of institutional path dependence.

Claim 2. (Existence of Path Dependence) *Any set of games that contains at least one susceptible game and two games with distinct 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, there exists one ordering of the games such that all susceptible games produce outcome A and another ordering in which all produce outcome B.*

The existence of a susceptible region enables path dependence. However, a larger susceptible region does not necessarily imply greater path dependence; the size of the susceptible region depends both on the amount of cultural sway and the historical context. One historical context could have a larger susceptible region but include more previous games. These previous games can restrict path dependence. We make that intuition formal in the next claim.

Claim 3. (Greater Susceptibility Need Not Imply Greater Path Dependence) *There exist contexts Ω and $\hat{\Omega}$ with the same threshold such that the susceptible region for Ω contains the susceptible region for $\hat{\Omega}$, but that context Ω does not exhibit greater path dependence.*

It does follow that a larger susceptible region implies greater path dependence if there has existed at least one outcome of each type in both contexts.

Claim 4. (Distinct Outcomes and Path Dependence) *If two historical contexts with the same threshold each include one outcome of each type, then a larger susceptible region implies greater path dependence.*

A straightforward corollary of this claim is that choosing an institution with clearer incentives—i.e a θ further from the threshold—produces greater future path dependence because it makes the susceptible region larger.

Corollary 3. *Given any game in an historical context, clearer incentives, i.e. payoffs further from the threshold, increase path dependence.*

We have shown how the degree of path dependence is captured by cultural sway provided that both outcomes have occurred. As cultural sway becomes dominant (formally, in the limit as γ approaches one), the susceptible region can converge to the entire space. It follows that the strategy played in the first game will be played in all subsequent games. This implies sensitivity to the initial game, and *not path dependence.*

We can measure initial game dependence as the probability that a given future game has the same outcome as the first game given a random sequence of subsequence games. The next claim states that the extent of initial game dependence strictly increases in the amount of cultural sway.

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

The previous claim describes outcomes for γ near one. The same effect holds for less cultural sway as well. In Figure 5, we show results from 1000 simulations of our Tradition/Innovation Game. We plot the number of times that the final threshold lies on the same side of the efficiency cutpoint as the initial game against the number of times that it

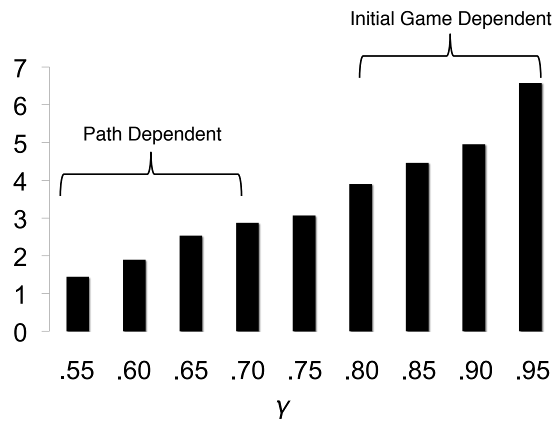

Figure 5: Odds Ratio of Threshold in Direction of Initial Game in Tradition Game after 1000 Epochs.

was not. This is the odds ratio that the initial game determines all subsequent outcomes. 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 determines a substantial majority of subsequent outcomes. We might more accurately describe those cases as initial game dependent.

These calculations demonstrate that if the outcome depends on the path, then both outcomes must remain possible. When cultural sway is large, the initial game determines behavior in nearly all future games.

Efficient Paths

We now derive necessary and sufficient conditions for an efficient sequence of games to exist and show how to construct such sequences. We restrict attention to sets of games that include at least one game in which outcome A is efficient and one game in which outcome B is efficient. We also require that at least one game lies in the immune region for one outcome (without loss of generality, we use B). Without that assumption, all games would produce

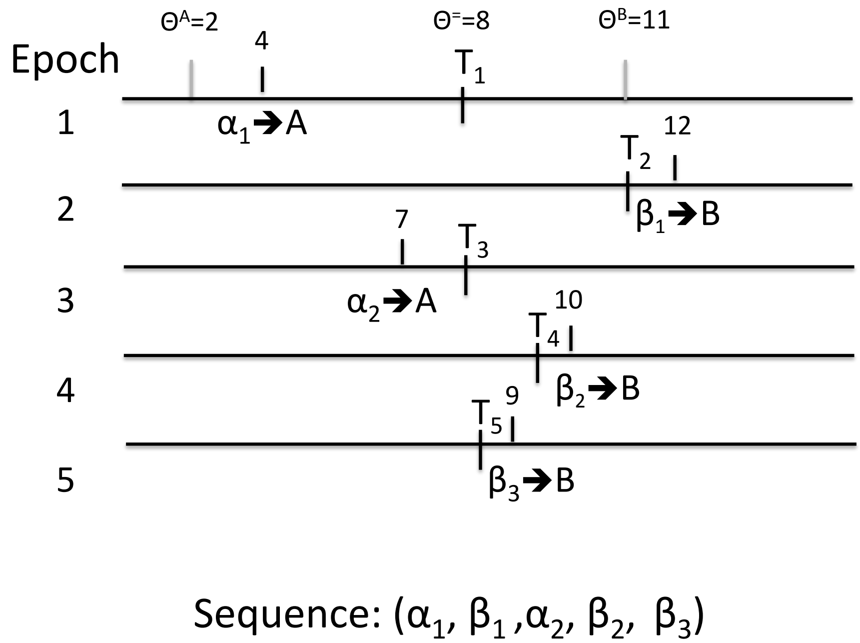

Figure 6: An Efficient Sequence of Games

the same outcome.

We first show that placing games with stronger incentives earlier in the sequence weakly increases the number of games with efficient outcomes. To state the claim, we introduce some notation. We can relabel the games corresponding to their efficient outcome and then index them based on the strength of the incentives they create. To be precise, given any set of games $\{\theta_1, \theta_2, \ldots, \theta_k\}$ we label and index the θ's as follows: For those games in which outcome A is efficient ($\theta < \theta^=$),we assign α labels. We then arrange the α's in increasing order. For those games in which outcome B is efficient, we assign β labels, and arrange the β's in decreasing order. We therefore have relabeled and reindexed the games so that $\{\alpha_1, \alpha_2, \ldots, \alpha_{k_\alpha}, \beta_{k_\beta}, \ldots, \beta_2, \beta_1\}$ where $\alpha_j < \alpha_{j+1}$, $\beta_i > \beta_{i+1}$.

Two principles underlie the construction of efficient sequences. First, games with lower indices, i.e. those with stronger incentives should be introduced earlier. Second, outcomes of both types should be alternated to some extent. This can be seen through an example in which we alter the sequencing of a common set of games. Assume that payoffs and cultural sway are such that the efficiency cut point equals eight ($\theta^= = 8$) and the susceptible region is bounded by two and eleven ($\theta^A = 2$, and $\theta^B = 11$) as shown in Figure 6. Finally let the set of games be $\{4, 7, 9, 10, 12\}$.

We first sequence the games according to their strengths of incentives alternating between games that have A as the efficient outcome and games that have B as the efficient outcome. This produces the sequence $(4, 12, 7, 10, 9)$. Refer again to Figure 6. As each game is introduced, the threshold (denoted by T_t) moves in the direction of the game just introduced. By construction, each subsequent game has sufficiently strong incentives that it lies on the appropriate side of the threshold. For example, game α_2 which has a payoff parameter of seven lies to the left of the threshold T_3 which equals eight.

We next consider an alternative sequence $(4, 12, 10, 9, 7)$ that does arrange the games by strength of incentives but includes all of the games with B as the efficient outcome before the second game that has A as the efficient outcome. This sequence violates the second principle. As can be seen in Figure 7, by the time that the game α_2 is introduced in epoch five, the threshold T_5 has fallen below seven, so the game now produces an inefficient outcome. Had the game been placed earlier in the sequence the outcome would have been efficient.

To make these intuitions more formal, we first state a claim establishing the benefits of ordering games by strength of incentives, a method of sequencing we call *incentive based incrementalism.* The claim states that given any sequence of games that produces efficient outcomes in every game, switching the order of the games so that those with stronger incentives (lower indices) are introduced earlier will maintain efficiency in at least as many future games.

Claim 6. Incentive Based Incrementalism *Given any set of games labeled* $\alpha_1 < \alpha_2 \cdots < \alpha_{k_\alpha} < \theta^= < \beta_{k_\beta} < \ldots \beta_2 < \beta_1$*, any game sequence in which there exists integers* j *and* j' *such that* $j > j'$ *and* α_j *(resp.* β_j*) appears prior to* $\alpha_{j'}$ *(resp.* $\beta_{j'}$*) produces inefficient outcomes in at least as many games as an alternative sequence in which game* α_j *appears before* $\alpha_{j'}$

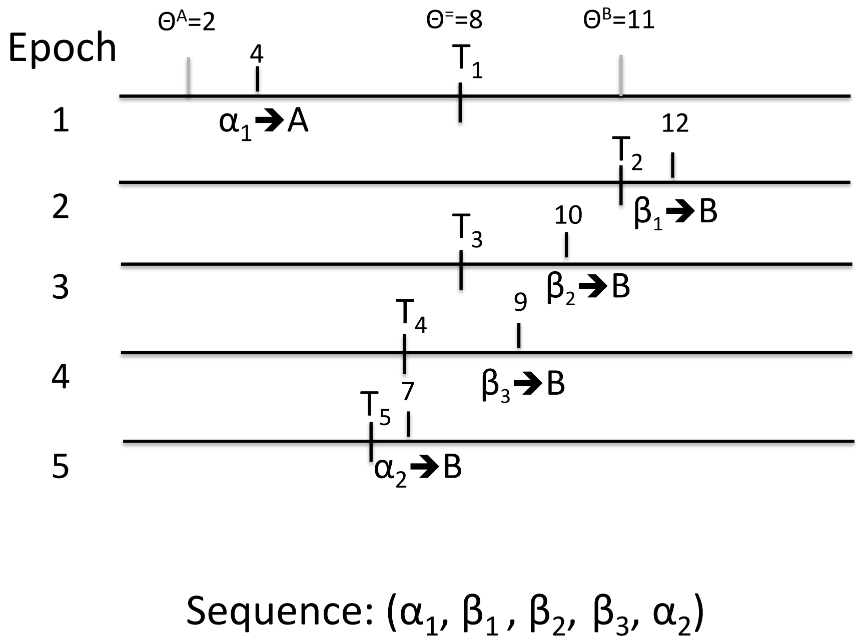

Figure 7: An Inefficient Sequence of Games

(resp β_i appears before $\beta_{j'}$).

In light of this claim, we hereafter assume games are introduced by increasing indices and that α_1 is the first game introduced so that the sequence produces an outcome A. Assume next an outcome of A in game α_{j-1}. We can then define the *immunity score* of game α_j to be the number of type β games that can be introduced (starting from the most extreme) yet still obtain an efficient outcome in game α_j game. To state this formally, the immunity score equals the largest number of β games that can be introduced prior to α_j such that those games all produce outcome B, yet game α_j still produces outcome A.

Given a game α_j (resp β_i), and a set of games $\{\alpha_1, ..\alpha_R, \beta_M ... \beta_1\}$, the **immunity score** *for α_j (resp β_j) is defined as follows:*

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

From this definition, games with large immunity scores will be less susceptible to the sequence of games. At one extreme, a game in the immunity region of outcome A (resp. B) has an immunity score equal to k_β (resp. k_α). At the other extreme, a game with an immunity score of zero must be introduced prior to any game that produces the other outcome.[14]

Our next claim relates the immunity scores to the possibility of an efficient sequence of games. Assume that $k_\alpha = k_\beta$, that is there are equal numbers of games with A and B as efficient outcomes. The claim gives a sufficient condition for the alternating sequence of games $(\alpha_1, \beta_1, \alpha_2, \beta_2, ...\alpha_{k_\alpha}, \beta_{k_\beta})$ to be an efficient sequence.

Claim 7. *(***Efficient Alternation***) Given a set of games with an equal number of efficient A outcomes and B outcomes, if $I(\alpha_j) \geq j$ for all j and $I(\beta_i) > i$ for all i, then the alternating sequence of games produces efficient outcomes in every game.*

The proof of the claim is straightforward. If the games are introduced in the order α_1, β_1, α_2, β_2 and so on, then by the construction of the immunity score, each game produces the efficient outcome. The alternating sequence will fail to be efficient if any game has an immunity score less than its index. For example, suppose that game α_5, which has an index equal to five, has an immunity score of three. This low immunity score means that only β_1, β_2, and β_3, can be introduced prior to α_5 yet still have game α_5 produce outcome A.[15] This implies that if the games are introduced using the alternating sequence, the outcome

[14] The immunity score obviously depends on the size of the behavioral spillover γ.

[15] This would mean that game β_4 is closer to game α_5 than is game α_4.

in game α_5 would be B.

Violation of the inequality in the previous claim does not imply that an efficient sequence cannot exist. If game β_4 has an immunity score larger than five, then game α_5 could be introduced prior to game β_4, and each game would still produce an efficient outcome. The following claim which gives necessary and sufficient conditions for the existence of an efficient sequence. We refer to the procedure of incrementally weakening incentives from each direction as *multi-directional incrementalism*.

Claim 8. (Efficiency and Multi-Directional Incrementalism) *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.* $j > I(\beta_i)$*,* $I(\alpha_j) \geq i$*.*

As stated in the next corollary, if a set of games does *not* permit an efficient sequencing, then an efficient sequence can be created by introducing new, more extreme games early.

Corollary 4. *Given a set of games for which no efficient sequence of games exist. An efficient sequence can be created by adding games to the set that have more extreme payoffs than the games that do not produce efficient outcomes.*

The theoretical results reveal a benefit to placing games with higher immunity earlier in a sequence. Societies that early in their history introduce institutions that produce diverse behaviors may be more likely to sustain that diversity. They can leverage that diversity to produce efficient outcomes in future games. The corollary suggests a lesson for reform: if you cannot attain an efficient outcome in the game you want, construct a new game with stronger incentives first.

Endogenous Institutional Change

Our framework connects to the *quasi-parameter* framework introduced by Greif and Laitin (2004). They describe a process of endogenous institutional change where game play produces feedbacks that change the payoff structure within an existing game. They refer to the changing payoff values as quasi-parameters. To translate their quasi-parameter to our model, one can consider incremental adjustments to the θ's of an existing game as the equivalent of new games being introduced. As the θ of an existing game changes, equilibrium behavior can be reinforced or become more fragile depending on the change in payoffs. A change in payoffs could degrade an equilibrium behavior if it makes that behavior inefficient.[16]

Institutional drift—the method of change in a quasi-parameter framework—implies costly transitions. A reinforcing quasi-parameter has no effect on efficiency. The equilibrium outcome was efficient and remains so. Degrading quasi-parameters are another matter. Suppose that initially, an institution had an efficient outcome A but that θ increases and crosses the efficiency cutpoint. Outcome B becomes efficient. Our model provides a framework for evaluating the transition. Behavior would not change—remaining inefficient—until 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 produces behavioral change only when entering 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

[16] If the change in parameter is sufficiently large to destroy the current equilibrium, then in Grief and Laitin, behavior would adjust immediately. If it makes the current equilibrium inefficient, then to model degradation, one would need to model explicitly how behavior changes as a function of that payoff change.

with each change in θ.

In Grief and Laitin's model, the quasi-parameter changes endogenously whereas we assume a sequence of exogenously chosen distinct institutions. In Grief and Laitin, behavior must change within the same institution. In our model, behavior is chosen in a new, similar institution. One might expect that cultural sway, i.e. behavioral stickiness, would be larger for endogenous changes to an existing institution than for a new and similar institution and that endogenous degradation would produce even more inefficiency. Within our model, behavioral stickiness would produces maximal inefficiency as the equilibrium only changes once the quasi-parameter drifts into the immune region. Our model suggests a solution to redress the cost of institutional transition through drift: speed up the degradation through a large change in the quasi-parameter that moves the game into the immune region for the efficient behavior, or, if appropriate, introduce a new, more extreme institution to germinate behavioral shift in the existing, drifting institution.

4 Results for General Classes of Games

We now extend our framework to cover all two by two symmetric games. Within this more general class of games, we show that when cultural sway is large, institutions that weakly punishments for deviation are more likely to produce efficient outcomes. Increasing rewards for choosing the correct strategy also improves the likelihood of efficient outcomes but not as effectively as weakening the punishment for failing to coordinate on the efficient equilibrium.

Our analysis relies on the following parameterization of two by two symmetric games:

	A	B
A	ω, ω	ρ, ν
B	ν, ρ	$0, 0$

The parameters ω, ν, and ρ can take any real value. This class of games admits three types of efficient equilibria: A and B as before, and a third in which players alternate between A and B that we denote by S. The efficient regions for each of the three strategies are shown in Figure 8.

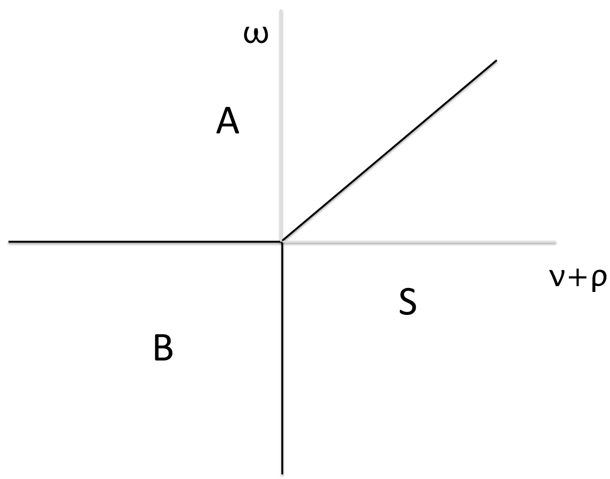

Figure 8: Efficient Regions For The Three Strategies in 2 x 2 Games

To analyze this more general class of games, we must modify our previous construction in two ways. First, we need to define a distance or similarity measure between games. We could use Euclidean distance over payoffs or a lexicographic measure in which a game is closer to games that have the same efficient equilibrium and then base distance on the Euclidean metric. The results that follow do not depend on the distance metric used; only that it is well defined.

Second, we must characterize off-the-equilibrium play, i.e. punishment strategies. Following convention, we assume punishment relies on the minmax strategy. 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. To avoid complications, we assume that within an epoch a game is repeatedly infinitely and that discounting is sufficiently low so that we can rely on average payoffs.

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 in this second game. Let M denote the minmax payoff. Assuming infinitely repeated play 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:

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

The efficient outcome will be achieved 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. Given 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* *infinitely 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) choose a game so that the nearest game has the same efficient equilibrium, (2) increase payoffs to the efficient equilibrium, or (3) increase the minmax payoff. The third route is the most powerful. If a new institution creates large punishments (a small minmax payoff) then the cost of overcoming cultural sway will be high. Mild punishments enable the efficient behavior to take hold.

5 Discussion

Whether implementing a new law, managing a transition—possible on a grand scale such as in a transition to democracy—or introducing policies to achieve more targeted goals like reducing obesity, the order that laws and institutions are introduced can matter, as scholars of development have long noted. Conflicting interpretations of the empirical evidence create an opportunity for foundational models of institutional sequencing to unpack the logic of when and how institutional context and, more generally, culture matter.

This need for models is reinforced by recent influential studies that establish 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). In our framework, we capture culture as a behavioral consistency across institutional domains, and generate results about how culture modeled in this way will affect institutional performance.

A model should explicate conditions for common intuitions to hold and fail, it should produce more subtle, and sometimes unexpected results, and it should enable one to ask new questions. Our model does all three. First, we've 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 even when spillovers are mild, and the level of inefficiency caused by this path dependence correlates 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.

Second, the model reveals less intuitive comparative statics. As cultural sway increases, the susceptible regions increase until 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. Under these conditions the first institution has an enormous effect on the agents' responses to subsequent institutions.

Finally, we derive rules for the optimal sequencing and design of institutions. We find that the key to efficiency is diversity of institutions and behaviors. Optimal sequences start from diverse extremes, creating incentives to generate distinct behaviors, and eventually introduce institutions where outcomes are more contingent on the past. Thus, the way to reduce *realized* path dependence is to keep its *potential* alive for as long as possible, by creating incentives for diverse behaviors early.

Our results challenge gradualism because it reduces diversity. When institutions evolve incrementally, behaviors become reinforced rather than adapting. The idea that optimal sequences of choices should maintain options, although new to political science, can be found in slightly different form in artificial intelligence. The best game-playing algorithms keep strategies open (Gelly et al 2012). We find that the best institutional sequences should do the same.

Regarding optimal design, we find that strong negative consequences are counterproductive: they reduce the likelihood of efficient outcomes by raising the cost of experimenting. This result runs counter to standard mechanism design logic that one should choose institutions with dominant strategies (Page 2012).

To conclude, our framework enables us to explore the implications of cultural sway and behavioral spillovers. We identify the necessity of taking into account behavioral repertoires when constructing or sequentially introducing institutions. Our analysis extends the formal literature on institutions by including the cultural and behavioral effects. We advocate proceeding in this new direction with vigor and caution. Models of microprocesses used to explain macrophenomenon inevitably fail to capture important aspects of the environment. These gaps limit our ability to draw inferences about the real world, to construct accurate hypotheses, and to design effective institutions. By filling those theoretical gaps with micro level data, formal institutional analysis can begin to bridge two literatures that rarely communicate: the stark theoretical models that isolate and identify informational structures and incentive effects, and the rich, comparative case studies that elucidate context.

References

Acemoglu, Daron and James Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty.* New York: Crown.

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." *QJE* 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.

Bednar, Jenna, Andrea Jones-Rooy, and Scott E. Page. 2015. "Choosing a Future based on the Past: Institutions, Behavior, and Path Dependence." *European Journal of Political Economy* forthcoming.

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

Boix, Carles. 2003. *Democracy and Redistribution.* Cambridge, UK: Cambridge University Press.

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.

Finnemore, Martha and Kathryn Sikkink. 1998. "International Norm Dynamics and Political Change." *International Organizations* 52(4):887–917.

Gelly, Sylvain, Marc Schoenauer, Michle Sebag, Olivier Teytaud, Levente Kocsis, David Silver, Csaba Szepesvri (2012). "The Grand Challenge of Computer Go: Monte Carlo Tree Search and Extensions." *Communications of the ACM,* Vol. 55, No. 3

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.

Jackson, Matthew O. and Yiqing Xing. 2014. "Culture-dependent Strategies in Coordination Games." *Proceedings of the National Academy of Sciences* vol 111,

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.

Page, Scott E. 2012. "A Complexity Perspective on Institutional Design." *Politics, Philosophy and Economics* 11: 5-25.

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

Proof of Claim 1: Let $\pi_i(\theta)$ denote the payoff if both individuals choose strategy i and 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 the immune region is empty, the result follows immediately. Assume an immune region for strategy A. The boundary of the immune region $\theta^A(\gamma)$ satisfies the following equation:

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

Simplifying gives:

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

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

Proof of 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 remains at θ^B. Therefore, assume $\theta_2 < \theta^B$, so that $T_3 = \theta^B$. 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 the largest θ_k for $k < t$ that produces outcome A and 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 it lies in the susceptible region. Otherwise, the threshold equals whichever of θ^A or θ^B is closest to that average.

Proof of Claim 2: Let θ^S denote a 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.

Proof of 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 assume that all three inequalities hold through time t and show that they then hold for time $t+1$. We consider three cases:

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 induction, $T_t(\gamma) \leq T_t(\hat{\gamma})$, therefore the outcome in the game θ_t is A for both spillover rates. Furthermore, $\psi_t^B(\gamma) = \psi_{t+1}^B(\gamma)$ and $\psi_t^B(\hat{\gamma}) = \psi_{t+1}^B(\hat{\gamma})$ so (iii) holds. To prove (ii) holds, by assumption $\psi_{t+1}^A(\gamma) = \theta_{t+1}$. There exist two possibilities: 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 prove (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 induction, $\psi_t^B\gamma \geq \psi_t^B(\hat{\gamma})$ 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 $\theta_{t+1} < T_t(\hat{\gamma})$. The outcome is B for γ and A for $\hat{\gamma}$. All three inequalities hold trivially. 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.

Path Dependence (Formal Def'n): Define $\Omega = \{\gamma, (g_1, g_2, g_3, ..g_k)\}$ and $\hat{\Omega} = \{\hat{\gamma}, (\hat{g}_1, \hat{g}_2, \hat{g}_3, ..\hat{g}_{k'})\}$ to be **outcome equivalent** if and only if for any game played next, the same outcome is produced in both contexts, $g \in G$, $\Omega(g) = \hat{\Omega}(g)$.

Given Ω and $\hat{\Omega}$ that are outcome equivalent. We say that Ω exhibits **greater path dependence** if and only if for any game, the set of sequences of future games that changes the outcome in game g in context Ω strictly contains the set of sequences of future games that change the outcome in context $\hat{\Omega}$.

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

Proof of Claim 3: The proof uses payoffs from the traditional and innovative strategies game and relies on a counter example. Assume context $\Omega = \{0.8, (1, 15)\}$ and context $\hat{\Omega} = \{0.75, (\emptyset)\}$, where $\emptyset$ denotes the empty set. Initially, $T = \hat{T} = 8$. The susceptible region in context Ω contains the susceptible region for context $\hat{\Omega}$. Now, consider the game $\theta = 1$. In Ω, the threshold does not change. However, in context $\hat{\Omega}$, $\hat{T}$ move to 10. This means that the sequence $(1, 9)$ would produce an inefficient outcome in $\hat{\Omega}$ but not in Ω. Therefore, Ω cannot be more path dependent.

Proof of Claim 4: Let $T = \hat{T}$ equal the common threshold in both contexts. Let θ^a equal the largest θ_k in context Ω that produces outcome A and θ^b equal the smallest θ_k in context Ω that produces outcome B. Define $\hat{\theta}^a$ and $\hat{\theta}^b$ similarly for context $\hat{\Omega}$. 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 $\theta < T$ so the outcome equals A. We first state a lemma.

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 $\hat{\Omega}$.*

First, 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 and 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 produce different outcomes is larger in context Ω than in context $\hat{\Omega}$. Therefore, after one game has been added, context Ω produces more path dependence than context $\hat{\Omega}$. Note that given any sequence of future games, 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 $\hat{T}$*

We can assume that the new game produces outcome A, i.e. $\theta < T$. 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 $\theta^a < \hat{\theta^a}$.

Proof of Claim 5: By Claim 1, the size of the initial susceptible region weakly increases in γ. To show that initial path dependence strictly increases, we must show first that for any sequence of future games $(\theta_1, \theta_2, ...\theta_k)$, 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 sequence of future games that produces a different outcome given γ but not given $\hat{\gamma}$. It suffices to consider cases where the first outcome is A. In any sequence of future games all outcomes will be 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 sequence of future games that produces an outcome of B for some game under γ but not under $\hat{\gamma}$, consider the single game sequence of future games, $\theta_2 \in (\theta^B(\hat{\gamma}), \theta^B(\gamma))$. 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].

Proof of Claim 6: Assume that there exists an $i < j$ such game α_i appears before α_i. Let j be the last α_j in the sequence for which this occurs. It follows that the next α game in the sequence, α_i, will satisfy the condition $i < j$. Thus, any games that appear between α_j and α_i will be β games. Let t denote the epoch in which game α_j appears. Note that if game α_j produces outcome A, then so must game α_i. Therefore, there exist three possible pairs of outcomes in the original sequence.

Case 1: Both α_j and α_i produce outcome A: Construct a new sequence by moving game α_i before game α_j leaving all other games unchanged. The threshold in epoch t exceeds α_j in both sequences, therefore in the new sequence, game α_i produces outcome A. The threshold in the new sequence in epoch $t+1$ is weakly larger than the threshold in epoch t, so game α_j produces outcome A. The thresholds for all subsequent games are unchanged from the original sequence.

Case 2: Both α_j and α_i produce outcome B: After epoch t, the threshold is less than α_j in the original sequence, so β games that follow game α_j produce outcome β. Construct a new sequence, by moving game α_j after game α_i and moving all β games that occur after α_j before α_i. All the β games moved still produce outcome B because in the original sequence, the threshold at t had to be less than the efficiency cutpoint, $\theta^=$. It remains to consider the

α games. If game α_i produces outcome B, then game α_j faces a weakly lower threshold than in the original sequence and still produces outcome B. Further, any α games that follow have unchanged thresholds. If in the new sequence game α_i produces outcome A, but game α_j produces outcome B, then once again, the thresholds for all subsequent α games will be unchanged. (Note that the new sequence is more efficient because it produces an efficient outcome in game α_i.) If both games α_i and α_j produce outcome A in the new sequence, then by the proof of lemma 1, the thresholds for all games that occur after α_j are larger than α_j. Previously, those thresholds had been less than α_j, therefore, all subsequent games are more likely to produce efficient outcomes.

Case 3: Game α_j produces outcome B and game α_i produces outcome A: Construct the same sequence as in Case 2: move β games that occur after epoch t ahead of game α_j and then switch the order of games α_i and α_j. In the original sequence, game α_i produces outcome A. If game α_i is immune it produces outcome A. Assume not. The threshold faced by game α_i in the original sequence was the midpoint of the smallest α that produced outcome B (possibly game α_j) and the largest α that produced outcome A. In the new sequence, α_j appears after α_i so the threshold when game α_i appears has a weakly larger value than in the original sequence. Therefore, game α_i produces outcome A. If game α_j produces outcome B, then the thresholds for all subsequent games are unchanged in the two sequences. If game α_j produces outcome A, the thresholds for all subsequent games will be greater than α_j. In the original sequence, the thresholds for those games were less than α_j, so those games are more likely to produce efficient outcomes.

Proof of Claim 8: 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.

Proof of Claim 7: We first prove sufficiency. Suppose no games exceed balanced sequencing. It suffices to consider the case where $R < M$ so that 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 occur earlier in 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 balanced sequencing is violated. Let $I(\alpha_{j'})$ be the first α's that exceeds balanced sequencing and $I(\beta_{i'})$ be the first β's that does. 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.

By symmetry, assume that $I(\alpha_{j'}) < I(\beta_{i'})$. Games can be added by the following algorithm.

Step 1: Up to game $I(\alpha_{j'})$ use the alternating sequence.

Step 2: Add all α games up to $\alpha_{j'}$.

Step 3: If no remaining games exceed balanced sequencing add them according to the alternating sequence. If not, choose the unique game with the smallest index that exceeds balanced sequencing and go to Step 1.

This algorithm produces efficient outcomes in all games. By assumption, efficient outcomes exist for all games with indices less than j' in both sequences and for games α_j through $\alpha_{j'}$. Suppose that in Step 3, no remaining games exceed balanced sequencing. 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. If later games exceed balanced sequencing, the result follows by an identical logic.

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. By symmetry, assume $\hat{j} \leq \hat{i}$. Given our assumption that he conditions are violated, there exists a β_i s.t. $i > I(\alpha_j)$ with $I(\beta_i) < j$ Suppose that $\alpha_{\hat{j}}$ comes before β_i. By assumption, $I(\beta_i) < \hat{j}$, which implies that β_i produces an inefficient outcome. Alternatively, suppose that β_i occurs before α_j. By assumption $i > I(\alpha_j)$, then α_j produces an inefficient outcome.

Proof of Corollary 4: Assume game β_2 is the first game that produces an inefficient outcome. That is, it is closer to game α_2 than it is to game β_1. Suppose that game β_2 is closer to α_1 than to β_2. Let $\beta_2^1 = \frac{1}{2}(\beta_1 + \alpha_2) + \epsilon_1$ for some small $\epsilon_1 > 0$. If β_2 is closer to β_2^1, the proof is complete. If not construct β_2^2 that is closer to β_2^1 than it is to α_2 by setting $\beta_2^2 = \frac{1}{2}(\beta_2^1 + \alpha_2) + \epsilon_2$ for some small $\epsilon_2 > 0$. By construction, the outcome in game β_2^2 is B. One can construct a sequence of β_2^n similarly such that outcome in each game is B. If the ϵ_n converge to zero then the β_2^n converge to α_2 so for some m, β_2 is closer to β_2^m than it is to α_2 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.