Modeling the Institutional Matrix:
Norms, Culture, and Robust Design[1]

Jenna Bednar

April 2018

ABSTRACT

An institution's ability to shape behavior is affected by other institutions that are also active---and the legacy of institutions that preceded it.   What North described as an "institutional matrix" is a research domain of nearly untapped potential, but it requires new methodological approaches.   In this article I suggest one approach to modeling the connection between institutions, Games +, which considers multiple games simultaneously.   Connections between games take shape as behavioral spillovers, *the way that people employ familiar behaviors as an initial approach to a novel situation.   Using a Games + approach is a tool to provide insight into two fields of institutional research: the relationship between culture and institutional performance, and robust institutional design.*

Modeling institutions as games has yielded a powerful collection of insights connecting patterns in outcomes to institutional features.   Most models consider institutions in isolation.   This prevailing paradigm has created four lingering challenges: (1) identical, or at least similar institutions can produce different outcomes; (2) societies develop identifiable behavioral signatures, or cultures, consistent across problem spaces; (3) culture influences institutional performance and also emerges through the interactions mediated by those same institutions; (4) phenomena such as robustness are properties of ensembles of institutions.

   To make advances in tackling these challenges, one might make a model more realistic by embracing the full institutional context, what North (1994) described as an "institutional matrix." Nearly 25 years

---

later, the field is just beginning to build theories and applications that flesh out how institutions might be connected and how they may aid or impede one another's performance.

In this paper I describe a method to model multiple institutions explicitly and provide a method to address the four challenges, opening up new avenues for research. The framework takes a "Games +" approach: agents play multiple games simultaneously. Using it, one can trace the effect of spillovers between institutions—formal and informal—as each shapes behavior. I describe it, and then suggest how it might help answer the challenges listed above. The model may explain how social norms emerge and change and how institutions and culture coevolve. I then describe a distinct research domain, robust institutional design, where the goal is to design a set of institutions that maintain their functionality in the face of perturbations. These research questions rise up in a broad array of applications, from economic development to constitutional interpretation to treaty effectiveness and enforcement.

Institutions as Games +

We can define an institution as a game form that specifies information, agents, roles, beliefs and incentives.[2] Most early papers analyze a single game in isolation. Exceptions that account for the multiplicity of contexts do so in two ways. A first approach connects payoffs and/or actions across games. For example, Putnam's two-level games (1988), Tsebelis' nested games (1990), and Fearon's domestic audience costs (1994), assume that actions taken in one strategic realm have implications for another. Most recently, Slaughter (2017) develops this insight, coupled with network theory, to describe the interconnectedness of strategic games in a globalized world. A second approach allows reputation effects to indirectly link games through time and beliefs (eg Axelrod and Keohane 1985, Drezner 1999, Downs and Jones 2002).

In this section I'll suggest an alternative approach to analyzing institutions simultaneously by connecting institutions through the behaviors that they produce.

Behavioral Spillovers

Institutions are designed to shape behavior. When agents learn a behavior in response to one game, that behavior becomes part of their

---

[2]To this list I'll add that institutions also have a purpose, often formally defined as a social welfare function.

repertoire. When confronted with a new game, that repertoire informs their initial behaviors. A *behavioral spillover* is when an agent uses a familiar tool in a new context; in solving new problems, humans naturally rummage first through their existing toolkit (Swidler 1986). "Games +" is a method of linking institutions through behavioral spillovers: each institution, or game, has effects that extend beyond the outcomes of a single game (Bednar and Page 2007, 2018).

For example, to solve the "problem" of how to dance, the tango is much more likely to emerge in societies where people greet one another with warm hugs and kisses on the cheek than in societies where friends shake hands. Relying on routine behaviors can explain suboptimal behavior: Societies who seal a trade by handshake backed by personal acquaintance have a harder time scaling up trade routes than those that conduct trade through contracts backed by formal legal systems. The Games + approach of focusing on behavioral spillovers has some analytical cousins, including reasoning by analogy (Winston 1980, Gentner et al 2001) or using "case-based decision theory" (Gilboa and Schmeidler 1995).[3] Cross cultural empirical studies provide evidence of analogizing in small-scale societies, where agents liken a new institution to one they play regularly (Henrich et al (2004).

Unlike sociological approaches, the Games + approach does not assume that these behaviors remain fixed. Agents learn, observing what others do, and copy better actions. The equilibrium behavior that emerges depends on the behaviors of others. A cooperative equilibrium will be more likely to emerge in a cooperate culture, e.g. one in which people's behavioral repertoires include cooperative behavior. Results include behavioral stickiness, different responses to identical institutions by different communities, and the emergence of coherent behavioral characteristics within a society, i.e. high trust and low trust. Outcomes that had been dismissed as suboptimal play—incompatible with rational choice analysis—can be explained in terms of behaviors that arose in similar domains (Bednar and Page 2007).

Linking Institutions

One of the key challenges in multiple game analysis is to determine what makes the games similar enough to be related in an agent's mind. The first step is to model what people understand about their choices.

---

[3] Reasoning by analogy is a fundamental approach to legal reasoning (Sunstein 1993, Brewer 1996).

Mental models (Denzau and North 1994) are an agent's way of reducing the complexity of their environment. In a single-institution analysis, Mantzavinos, North, and Shariq (2004) and Greif (2006) describe how shared mental models or belief systems may explain the formation and evolution of institutions. Gibbons, LiCalzi, and Warglien (2017) describe how the frames created by these categorizations can lead to suboptimal outcomes.

With a mental construct in hand, we can return to the problem of determining how agents perceive relatedness. Schema are a way of describing mental models across multiple institutions. An agent's use of a schema is akin to the cognitive science problem of categorizing, where agents draw inferences from a subset of games to partition the game space into categories. New games can be binned according to their similarity to other games.

The emergence of schema is undertheorized, and virtually no work attempts to explain the direction of flow in how institutions affect one another. In our work, we've used two approaches to measure similarity and to predict the direction of influence: game difficulty and payoff similarity. In experiments (Bednar et al 2012, Liu et al 2018), we gauge the difficulty of each game by the uncertainty (entropy) of players' actions. As hypothesized, behavior in simpler games spilled over into more difficult games. In mathematical models (Bednar and Page 2018), we relate games by the similarity of their payoff structure, with parameters in the payoffs supplying a measure of incremental change.

Informal institutions, Norms, and Culture

Formal institutions have written rules, but even black-and-white text requires interpretation to predict behavior, as anyone who has driven on the interstate in Texas knows: drivers take cues from drivers around them as they decide how far down to push the gas pedal. All formal rules have an element of social construction; people look to one another's behaviors to assess the likelihood of enforcement and to coordinate their behavior. Increasingly legal scholars turn to informal institutions, like norms, to explain how law shapes behaviors as wide-ranging as tax compliance (Posner 2000), the operation and scope of the shadow economy (Schneider and Enste 2013), and respect for human rights (Risse-Kappen et al 1999). The tight relationship between norms and law is evident in the work of legal philosophers like Habermas (1996) who argue that law is not just about what one should not do, but also prescribes what one ought to do. And legal theorists increasingly describe how constitutions,

as written sets of rules, are completed through informal institutions (Griffin 2016).

Informal institutions such as norms differ from formal institutions because they are socially monitored and enforced. Actors watch one another; norms are reinforced when actors conform as well as when actors punish one another for any violation of expectations. Social enforcement makes norms acutely sensitive to the social context. Cries of "hey, that's not right!" are subjective, where the subjective perception is a function of the objector's past experiences (Bicchieri 2016). Recalling the discussion above of mental models, Greif and Mokyr (2017) argue that cognitive rules are socially constructed. Agents learn from one another by observing one another's behaviors, whether through speech or action.

In short, norm monitoring and enforcement is a behavior, and it is susceptible to influence by behavioral spillovers. Spillovers may be stronger or weaker, affecting an institution's relative vulnerability or the force of the spillover. That in turn may depend on how broadly—across the society's institutions—the spillover reaches, or how deeply—within a society's population—the spillover is carried. When spillovers have broad reach, they may become more fixed.

Returning to the concept of a schema, spillovers may be lumped together and labeled by a philosophy, ideology, or set of principles, which in turn can serve as a heuristic for choosing actions. Behaving consistently with other aspects of one's guiding philosophy may be cognitively less taxing. If one is concerned with norm change, overcoming an enmeshed norm may be particularly challenging. Bicchieri's (2016) empirical work on norms and norm change reveals that norms backed by a religious creed are harder to change. In Bednar et al (2010), we examine the cognitive tension an agent faces between coordination and consistency: social agents want to align their behavior with others, but also face cognitive dissonance if that behavior is out of step with their other practices. The presence of two conformity-rewarding forces did not speed the rate of full convergence of the system; instead, we observed persistent diversity as well as the emergence of a cloud-like general "cultural" signature, much as we might have a general sense of how a French woman or Russian man might react to a situation, even as we recognize how different individual French women or Russian men can be.

Our ability to make predictive generalizations about members of particular communities comes from the coherence of cultures. Culture is more than the aggregation of otherwise unrelated norms. Culture

implies a congruity across gestures, symbols, art, dress, housing, and cuisine, and in the way that its adherents treat one another and treat outsiders. Building a model of culture that embraces this interrelatedness is a new and open project. The Games + approach enables us to model institutional relatedness, predict the direction of influence, and draw inferences from patterns of behavior to formulate expectations for how a member of a society might react to a novel situation.

Institutional Path Dependence

Thinking in terms of spillovers is a way to operationalize institutional path dependence postulated by Manzanvinos, North, and Shariq (2004). A Games + framework implies that institutions shape behavior, and behavioral spillovers, patterned as culture, shapes the performance of institutions and the design of future institutions. Therefore, there exist better and worse orderings of institutions.

In our framework, we introduce institutions sequentially to a society. Agents develop behaviors in response to initial institutions; under some conditions, early play shapes future choices. Some institutions are path-limiting, curtailing the flexibility of agents' future play (Bednar et al 2015). In Bednar and Page (2018), we build prescriptions for optimal sequencing based upon the lessons of path dependence: to avoid lock-in, introduce institutions with widely varying incentives early in the sequence. Doing so ensures the breadth of a populations' behavioral repertoire, minimizing the extent that it is trapped in suboptimal play. We also use this model to understand the ideal severity of sanctions.

To summarize this section, a Games + framework—analyzing multiple games simultaneously and the spillovers between them---can provide insight into analytical challenges like behavioral stickiness, suboptimal behavior, and varying institutional performance. It may suggest pathways for introducing institutions into a society, including optimal intermediate institutions that might prime the development of conducive behaviors. It offers one approach for studying the emergence and influence of culture. And it may provide a means to derive testable hypotheses regarding cultural dynamics, including an explanation for the apparent tendency of some societies to get stuck, when culture becomes absorbing, inhibiting innovation, while the culture of other societies seems to facilitate adaptation as the world changes. We believe that this systems perspective offers abundant research opportunities.

Robust Institutional Design

The second broad research domain, robust institutional design, requires a pivot in perspective. Robustness, the maintenance of functionality in response to disturbances and internal dynamics, differs from stability, which refers to a system's return to equilibrium. In line with the Games + approach, robustness often results from the interaction of a set of institutions. Traditional mechanism (institutional) design focuses on implementing a social choice function, that is, creating a game in which the Nash equilibrium is the desired outcome. In a robustness perspective, the goal is to design an institution that maintains its functionality in a changing world, where institutions balance the production of meaningful outputs and allocations against the search for new and better solutions. Technological changes, climactic events, international disturbances, and internal political and social dynamics all contribute to a changing environment within which a set of institutions must function. The goal is to design institutions that are *adaptively efficient*.

A robust institutional design frame requires a sharp definition of an institution's function and functional failure. Both are harder to do than they sound. What is the purpose of democracy? We might think that a democracy's purpose would be to maximize representation or provide for efficient allocation of public goods. How far from majority rule would a democratic system have to go to be considered a failure? In general, it is best to define function more narrowly.

In Bednar (2009), I evaluate the robustness of federal systems. The purpose of a federation is defined as a combination of security, economic growth, and representation. To tighten this sprawling function, I focus on the authority boundary that distinguishes the national and state jurisdictions, drawn with the intention of generating a governance structure that produces federalism's benefits. A system of safeguards—separation of powers, the judiciary, the party system, and public input—defend federalism's boundaries from regular threats of state and federal government overreach. But trickily, these safeguards must also be flexible, in order to allow the boundaries to respond to changes in the environment that make federalism's benefits more accessibly reached. Robust systems of institutions must exhibit flexibility yet produce good behavior. These contradictory incentives exemplify March's (1991) exploration/exploitation tradeoff: to what extent should an organization exploit current knowledge or search for new solutions?

Robust institutional design has four key properties: complementarity, redundancy, modularity, and diversity. The first three components contribute to a failsafe system; the last, with some help from

modularity, facilitates its adaptation. With complementarity, one institution's enforcement mechanism is bolstered by the presence of a secondary mechanism, provided by another institution: courts acting alone are toothless; executive police powers back the enforcement of judicial remedies. If the executive doubts the court, the court tends to move away from controversial decision-making and hews more closely to the public will (Friedman 2009). Redundancy refers to multiple pathways to functionality; in the federalism example, several safeguards may intervene in the event that another institution fails. It is important that the institutional enforcement mechanisms have different sources of failure; that is, they should not be reliant on identical information, agents, or incentives. Modularity allows for changes in one institution to have minimal effects on others. Experimentation by states in a federal system are one example of modularity, as is the separation of the executive and legislative branches of government. Finally, diversity injects new ideas. In a federal system, diversity comes from the modularity of different states as well as the multiple viewpoints incorporated into the variety of safeguards.

One key insight from robust design theory is that an ideal ensemble of institutions need not consist of perfect institutions (Bednar 2009, Vermeule 2011). If agents are mildly noncompliant, they explore the boundary, testing to see whether there is a better way to organize authority. What one institution sees as a transgression may well be a policy innovation. Institutional imperfections keep the door ajar so that diverse ideas might sneak through. Paradoxically, tighter (more consistent, coherent) sets of institutions, whether formal, like a totalitarian state, or informal, like a cultish culture, could be more brittle as they admit less diversity of thought, behavior, and norms.

Future models might well combine these two problem domains, culture and robustness, in the following way: if an enforcement mechanism relies on the public, the public may carry interpretations from one context to another. And, those interpretations could, at least in theory, subvert what had seemed a robust design. For example, as I write this article, some argue that United States' democracy is being tested. Democracy is supported by a variety of complementary and redundant institutional safeguards, including various ways that powers are separated. But behind the fragmentation of power and other formal institutional safeguards lies a public insistence on rule of law, a philosophy that collects together expected reactions to violations of putting the public good over private interests. If public commitment to rule of law wavers, if the people allow laws to be interpreted in any

which way, they will prove an insufficient backstop to the formal fragmentation of authority, and democracy could be vulnerable. Thus, harkening back to this essay's opening, similar institutions could produce a robust federation in one setting although not in another, but now we can test whether varying institutional performance is due to behavioral and normative spillovers.

Conclusion

In this brief article I've pointed scholars toward studying how institutions are connected---North's ``institutional matrix''---a domain that is begging for exploration. I've offered a method of analysis, Games +, that invokes behavioral spillovers as the connective tissue between institutions. I've described two broad research domains where one might apply an institutional matrix model: culture and robust design, each offering much room for methodological development and bursting with possible applications.

The upshot of this approach is that institutional analysis may require a third focus: behavior. Traditionally, in institutional analysis we've cared about the outcome or the process---asking, for example, whether the ends are just or the means are fair. Behavior has almost been incidental to our studies. In a Games + approach, we see that the behavior generated by an institution matters as well because that behavior has implications for the process or outcome of other institutions. By bringing behavior, and behavioral spillovers, into theoretical focus, institutional analysis may move farther from abstract blackboard theorizing, to building a science of real-world understanding.
.

References

Axelrod, Robert, and Robert O. Keohane. 1985. "Achieving cooperation under anarchy: Strategies and institutions." *World politics* 38(1):226–254.

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

Bednar, Jenna, Aaron Bramson, Andrea Jones-Rooy, and Scott E. Page. 2010. "Emergent Cultural Signatures and Persistent Diversity: A Model of Conformity and Consistency." *Rationality and Society* 22(4):407–44.

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* 40(B):312-322.

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 and Scott E. Page. 2018. "When Order Affects Performance: Culture, Behavioral Spillovers, and Institutional Path Dependence." *American Political Science Review* 112(1):82–98.

Bednar, Jenna, Yan Chen, Tracy 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.

Bicchieri, Cristina. 2016. *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Brewer, Scott. 1996. "Exemplary reasoning: Semantics, pragmatics, and the rational force of legal argument by analogy." *Harvard Law Review* (1996): 923-1028.

Denzau, Arthur T., and Douglass C. North. 1994. "Shared mental models: ideologies and institutions." *Kyklos* 47(1): 3–31.

Downs, George W., and Michael A. Jones. 2002. "Reputation, compliance, and international law." *The Journal of Legal Studies* 31(S1): S95-S114.

Drezner, Daniel W. 1999. *The sanctions paradox: Economic statecraft and international relations*. New York: Cambridge University Press.

Fearon, James D. 1994. "Domestic political audiences and the escalation of international disputes." *American Political Science Review* 88(3): 577–592.

Friedman, Barry. 2009. *The will of the people: how public opinion has influenced the Supreme Court and shaped the meaning of the Constitution*. Farrar, Straus and Giroux.

Gentner, Dedre, Keith James Holyoak, and Boicho N. Kokinov, eds. The analogical mind: Perspectives from cognitive science. MIT press, 2001.

Gibbons, Robert S. and LiCalzi, Marco and Warglien, Massimo, What Situation Is This? Coarse Cognition and Behavior Over a Space of Games (August 2017). Department of Management, Università Ca' Foscari Venezia Working Paper No. 2017 / 09.

Gilboa, Itzhak, and David Schmeidler. 1995. "Case-based decision theory." *The Quarterly Journal of Economics* 110.3: 605-639.

Greif, Avner. 2006. *Institutions and the path to the modern economy: Lessons from medieval trade*. Cambridge University Press.

Greif, Avner, and Joel Mokyr. 2017. "Cognitive rules, institutions, and economic growth: Douglass North and beyond." *Journal of Institutional Economics* 13.1: 25-52.

Griffin, Stephen M. 2016. "Understanding Informal Constitutional Change." Tulane University School of Law, Public Law and Legal Theory Working Paper Series, Working Paper No. 16-1.

Habermas, Jürgen. 1996. *Between facts and norms: Contributions to a discourse theory of law and democracy*. Cambridge, MA: MIT Press.

Henrich, Joseph Patrick, ed. *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*. Oxford University Press, 2004.

Liu, Tracy Xiao, Bednar, Jenna, Yan Chen, and Scott Page. Forthcoming 2018. "Directional Behavioral Spillover and Cognitive Load Effects in Multiple Repeated Games." *Experimental Economics*.

Mantzavinos, Chrysostomos, Douglass C. North, and Syed Shariq. 2004. "Learning, institutions, and economic performance." *Perspectives on politics* 2(1):75–84.

March, James G. 1991. "Exploration and exploitation in organizational learning." *Organization science* 2(1):71–87.

North, Douglass C. 1994. "Economic Performance Through Time.? American Economic Review 84 (3): 359? 68.

Posner, Eric A. "Law and social norms: The case of tax compliance." Virginia Law Review (2000): 1781-1819.

Putnam, Robert D. "Diplomacy and domestic politics: the logic of two-level games." International organization 42.3 (1988): 427-460.

Risse-Kappen, Thomas, Stephen C. Ropp, and Kathryn Sikkink, eds. The power of human rights: International norms and domestic change. Vol. 66. Cambridge University Press, 1999.

Schneider, Friedrich, and Dominik H. Enste. *The shadow economy: An international survey*. Cambridge University Press, 2013.

Slaughter, Anne-Marie. The chessboard and the web: Strategies of connection in a networked world. Yale University Press, 2017.

Swidler, Ann. "Culture in action: Symbols and strategies." American sociological review (1986): 273-286.

Sunstein, Cass R. "On analogical reasoning." Harvard Law Review 106.3 (1993): 741-791.

Tsebelis, George. Nested games: Rational choice in comparative politics. Vol. 18. Univ of California Press, 1990.

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

Winston, Patrick H. "Learning and reasoning by analogy." Communications of the ACM 23.12 (1980): 689-703.