Three Models of Institutional Incongruity: Multidimensionality, Networks and Culture

Jenna Bednar, Jean-Paul Faguet, and Scott E Page

May 27, 2025

Abstract

Institutional and policy reforms often produce unintended consequences that result in institutional incongruity. Using the lens of complex systems theory, we describe three potential sources of complexity-induced incongruity: (1) *policy multidimensionality*, (2) *network interactions*, and (3) *interdependent institutions and culture*. Policy multidimensionality can take two forms: (1.1) *interdependent policy effects*, a top-down phenomenon that arises when policies interact with each other directly; and (1.2) *population realignments*, a bottom-up mechanism in which individuals re-self-organize according to an additional dimension that suddenly becomes salient. We analyze these complexity effects using game-theoretic and agent-based models and offer two insights related to the origins of institutional incongruity: First, that real-world instances of incongruity depend on complexity. In fact, some complexity may be necessary for incongruity to arise. Second, real-world cases of incongruity typically involve multiple complex causes, such as when policy dimensions become interdependent through population realignments.

1 Introduction

The history of institutional reform reveals the complexity of policy making. Policies intended (or claimed) to be for one purpose often result in incongruous or even contradictory effects in other dimensions. Devolutions of political authority intended to improve efficiency and retain power result in political realignments in which the governing party loses power; criminal justice reforms to reduce arbitrary detentions and improve the administration of justice instead lead to huge increases in extralegal killings of suspects by the police; attempts to promote women's political power devolve into symbolic representation as elite men design in specific features that undermine gender equality; and language reforms intended to boost human capital instead seriously undermine it.

Analysis of specific cases demonstrates that institutional interventions and policies can produce incongruity for a variety of reasons. The papers in this Special Issue demonstrate how much we can learn from careful analysis of those cases. We can also learn from theory and models. In this paper, we sketch how a complex systems approach that focuses on the interplay between institutions and behaviors, beliefs, and norms can provide insight into how institutional and policy reforms produce effects that are incongruous, and perhaps even orthogonal, to stated intentions. Complex systems approaches can explicitly capture the interdependence and feedbacks that arise in political systems.

Complex system approaches are particularly suited to the study of incongruity because they widen the analytic aperture and consider how actions in one domain influence behaviors and outcomes in others (Arthur 2019, Miller and Page 2007). Cross-domain effects and feedbacks within domains are key causes of incongruity in political systems (and beyond). As we argue in the next section, incongruity demands complexity. If the world of policy reform were simple, reform processes would be straightfoward and reformers could not so easily fool citizens.

In what follows, we construct a collection of illustrative models to demonstrate insights

from complex systems theory relevant to the analysis of institutional incongruity. These models demonstrate three categories of complexity-induced incongruity: *policy multidimensionality*, *network interactions*, and *interdependent institutions and culture*. Policy multidimensionality can, in turn, take two forms: *interdependent policy effects*, a top-down phenomenon that arises when policies interact with each other directly; and *population realignments*, a bottom-up mechanism in which a self-organized set of individuals re-sort themselves when an additional dimension becomes salient. These models emphasize how incongruity arises from complexity.

But first, let us define *institutional incongruity*, which in turn is the result of *instrumental mismatch*. We follow this Special Issue's lead article in assuming that politicians pursue reform for the sake of both private and stated goals. Instrumental mismatch is defined as the degree to which the tools of policy reform – the specific actions and plans politicians deploy – align with the stated goals of a reform. "When stated and private goals are similar, instruments will tend to align. But when private goals are opposed to stated goals, or operate in a different dimension ('orthogonal'), the instruments of reform – the specific characteristics of reform design and implementation – will be poorly suited to stated goals. Instrumental mismatch will be high.

High mismatch will tend to produce institutions that are incongruous, meaning ill-suited to their core functions, including reaching stated goals. Incongruous institutions will tend to produce outcomes that are bad for society. Instruments that are well-matched to stated goals, by contrast, will tend to lead to congruous institutions, with ultimate outcomes that are good for society" (Faguet in this Special Issue, p.3).

We are far from the first to apply complex systems theory to the study of political economy. The field began with foundational work on political systems (Jervis 1997), institutional interdependencies (Bednar 2009), and micro-level models of cooperation and collaboration (Axelrod 1997). In brief, a complex systems approach to studying the political economy of

institutions differs from standard equilibrium and dynamic equilibrium approaches in four important respects. The first three differences relate to foundational assumptions about how political systems function and people act. First, a complex systems perspective assumes heterogeneous actors who rely on behavioral rules and interact within networks that provide local and global information. Rule-based actors are purposeful. They seek to achieve their objectives, but do not necessarily make optimal choices.

Second, a complex systems perspective focuses on multiple institutions operating as an ensemble rather than in isolation (Bednar and Page 2018). The effect of an institutional change or policy in one domain will be influenced by beliefs, behaviors, and norms in other institutions, as demonstrated in many cross-cultural experiments (Bowles 2018). Third, rule-based actors face either computational limits or costs of gathering and processing information, or both. This assumption, combined with the assumption that individuals interact within multiple institutions, can produce consistency in behaviors across institutional contexts as an emergent phenomenon.

Consistency emerges at both individual and group level because inconsistent behavior is cognitively costly. Consistency means predictability. When expected or socially enforced, these consistent behaviors become norms and can be interpreted as a component of 'culture' if one adopts a toolkit conception (Swidler, 1986). From this perspective, culture and institutions co-evolve. Institutions produce consistent behavioral repertoires, and those repertoires make some institutional types perform better than others (Bednar and Page 2018).

The final difference concerns the outcomes that emerge. Complex systems are not tied to the analysis of equilibrium and deviations from it. Here, the complex systems approach departs most markedly from social science theorizing based on rationality assumptions and equilibrium as a solution concept. Complex systems can produce four classes of behaviors: stasis (a more general alternative to equilibrium), periodicity, randomness, and complexity. Complexity is sometimes defined as somewhere between random and ordered states (Wolfram

2001).

Political economic systems exhibit all four classes of outcomes. Some dimensions converge to equilibrium while others produce complex patterns. For example, the production of most commodities follows relatively stable growth paths over time (an equilibrium rate of growth in output), but the time series of any commodity price, e.g. oil, wheat, or copper, is not in equilibrium. It is complex. By complex, we mean hard to predict.

The implications of systems with multiple classes of outcome types are profound. In equilibrium models of institutions, policy effects are evaluated using *comparative statics*. An institutional change or a new policy produces a smooth change in the existing equilibrium, or, in extreme cases, movement to a new equilibrium. By contrast, complex systems can tip suddenly from equilibrium into chaos, or from complex to static outcomes (Lamberson and Page 2012).

When analyzing institutional effects within complex systems, scholars do not only look at time series data. They also analyze network configurations. An institutional change, or a reform within a complex system, will produce multiple effects. For example, a development grant will imply both an infusion of resources and a rearrangement of networks.

Complex systems are notoriously robust to change. Ant colonies, ecosystems, brains, and economies all tend to react vigorously to interventions. In the face of shocks, invasions, or changes to environmental parameters, they maintain functionality through adaptation. This is also true of political systems (Bednar 2009). Policies often generate an initial impact that disappears in the long run. This often happens because changes to one part of the system cause changes in other parts that mitigate or even destroy the effect of the original reform. If long-term change does occur, it tends to be gradual and require adaptations in behaviors and networks.

And yet there are instances when complex systems break. Hence, scholars speak of complex systems as robust yet fragile. An example from our context might be a political

alignment holding together for a long period and then suddenly collapsing. When realignment happens, outcomes can be difficult to predict.

Institutional changes often produce S-shaped effects as behaviors spread through a population. Times series of impacts resemble those of the adoption of new technologies, or the spread of disease. Finally, complex systems can produce major restructurings. These are variously referred to as large events, punctuated equilibria, or realignments. Such restructurings can be *path dependent* (Arthur 1996).

To summarize, a complexity perspective views political economies as systems of diverse, adaptive actors operating within multiple institutions. These actors adopt (mostly) coherent sets of beliefs, behaviors, and norms across their institutional setting. To understand what occurs when one institution changes, we must consider changes to beliefs, behaviors, and norms, and how such changes influence other policy domains. Often, institutional changes create effects that ripple across other institutions indirectly through changes in culture and civic capacity. A complex systems approach allows us to model such effects, in marked contrast to equilibrium approaches that, for the most part, treat institutions in isolation (Page 2011).

Complexity leads naturally to incongruity, a condition that can make effective policy reform difficult. And some policies are simply more complex than others. For example, designing a national health care system for a modern economy is a far more complex challenge than redesigning a national postal service. Routing and building the Queen Elizabeth Line to expand the London Underground was a far more complex undertaking than establishing the ferry service from Southampton to the Isle of Wight. As a result of such complexity, even well-intentioned policies may turn out to be ineffective, or – worse – incongruous. The empirical record shows that governments' efforts to do complex things, like reduce inequality, increase economic growth, and build democratic systems, often fail or have modest effects, and that large government projects often cost far more than anticipated (Flyvbjerg and Gardner 2023).

In a complex political system, policies formulated within one domain generally impact others as well. For example, policies intended to increase economic growth typically produce a variety of effects on ecosystems, climate, immigration, inequality, crime, and often local communities, who may benefit from more resources or suffer congestion or relocation. These interdependencies mean that policies formulated with the best intentions can produce unexpected or incongruous impacts (Faguet in this SI). As an example, legislative malapportionment in Argentina, an attempt to tie big, rich Buenos Aires into the nascent federation on the basis of over-representing smaller provinces in Congress, led instead to Buenos Aires' domination of successive federal governments for over a century via a compensating executive malapportionment (Paniagua and Ricart-Huguet in this SI). Such interdependencies can differ greatly by place, implying that attempts in developing countries to build institutions similar to those that have performed well in developed countries face large, unseen hurdles (Boone and Faguet 1998, Easterly 2017, Faguet 2004).

This second effect has a surprising consequence. A political actor might justify a policy by emphasizing a particular outcome, when in fact their true intention is to make changes along some other political axis. Even when the former motive dominates debate, it is the latter motive that shapes the reform, meaning the details of its design and implementation, making this the more relevant lens through which to view policy outcomes. In this SI, for example, Laitin and Ramachandran describe how Rwanda switched its official language to English with the stated goal of better educational outcomes and greater integration into the global economy. Both goals had political support; how could they not? But the leaderships' true motive was to reduce the power of francophone Hutu elites in government and the broader society. So rather than train francophone teachers to speak and then teach English, they declared the switch and left teachers floundering in the classroom. The outcome was a sharp fall in educational attainment by the generation of children most affected by the change. Complexity in policy reform paved the way to these outcomes.

2 The Necessity of Complexity for Incongruity

We begin by postulating the necessity of complexity for policy incongruity. We make that argument in two steps. First, we distinguish complexity from ambiguity, uncertainty, and difficulty (Page 2008). An institutional or policy domain is *ambiguous* if it is difficult to interpret or open to multiple interpretations. It is *uncertain* if the mapping from policies to outcomes depends on hidden information or random effects. The outcome of an institutional change might turn out differently than expected due to incomplete or poor information or simple bad luck, not incongruity. *Difficult* problems are high-dimensional, with many interactions between features. Designing a well-functioning website for national health care is an example of a difficult problem. Failure to achieve an expected outcome on a difficult problem reflects a lack of state or organizational capacity. This is also not incongruity.

Incongruity requires multidimensionality and interdependence. A change to one institution affects another institution. The addition of a policy dimension results in a realignment of actors or the creation of local niches. Reforms designed around outcomes in one dimension affect outcomes in other dimensions, some of which may feed back into the first. Institutional rules and incentives interact with cultural norms, beliefs, and behaviors in ways that can amplify, attenuate, or nullify the intended outcomes of a reform. Complexity, which can be defined as phenomena that are difficult to predict, explain, engineer, or evolve, arises within systems. Complex systems consist of diverse, adaptive actors whose choices and actions interact in nontrivial ways to produce macro-level phenomena. Often these macrophenomena have emergent structures and functionalities. Economies are complex systems, as are political systems and ecosystems.

If incongruity requires complexity, what would it mean to adopt a complex systems approach to analyzing institutional and policy changes? We present one approach in figure 1. In a complex system, an institutional or policy change has *direct effects* on a *targeted dimension*. It also has *interdependent effects* on other dimensions, as well as impacts on behaviors,

beliefs, norms, and networks. Institutional changes can even influence expressed identities. This would be the case, for example, if a policy gave members of an ethnic community special rights or privileges. Its effects on other policy dimensions and on culture would not be simple *externalities*, that is, one-time shifts in equilibrium values. Rather, they would constitute *systems effects* that continuously influence the targeted dimension as well as other dimensions. *Systems effects* like these can ultimately undermine the original objectives of a reform. This is one kind of policy incongruity.

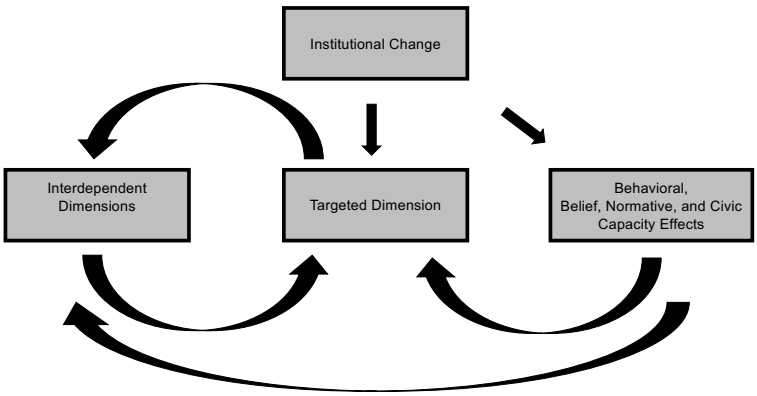


Figure 1: Direct and Interdependent Effects of Institutional Changes

Institutional change requires political support. Our analysis takes for granted the political support and state capacity to enact and execute a policy or institutional change. Political support generally implies popular support. A leader must convince others that a change will produce benefits, e.g. greater democratic voice, expanded rights and opportunities, reduced crime and corruption, or improved economic performance. We refer to these as the *targeted dimensions*.

The targeted dimension may not be the real motive for reform. Leaders' unstated motivations may operate in different dimensions. An infrastructure initiative may be proposed

as an engine of economic growth, but in reality serve as a means to shovel resources to a politician's supporters.

Alternatively, politicians' real motivation might operate in the same dimension as the stated goal, but in direct opposition to it. For example, Chauchard, Brulé and Heinze (in this SI) describe a reform that purported to increase the political power of Indian women in local government, but in fact was implemented so as to consolidate power in the hands of elite men. Complexity makes such misdirections easier.

Or consider Rwanda's shift to English language instruction, analyzed by Laitin and Ramachandran in this SI. The effects targeted by the policy were improvements in school performance and greater international economic integration. But Laitin and Ramachandran's evidence shows opposite outcomes: educational attainment fell as literacy worsened and fewer students progressed from primary into secondary school. The policy also produced interdependent effects in other dimensions. It disempowered – even denigrated – French-speaking elites. From a complex systems standpoint, changes of such magnitude would be expected to produce powerful adaptive responses. In this instance, those would take the form of behavioral feedbacks. In Rwanda, Anglophones gained power, jobs and social status when they filled positions vacated by French-speakers. Not only students, but people across the society faced new incentives to speak English. This reorganized the social hierarchy and reallocated power against the French-speaking majority.

Consider another example from this SI, this time from a developed country. In Hopkin's analysis, the drawn-out, immensely painful Brexit process was driven not by its purported targeted dimension – the UK's economic and political relationships with EU countries – but rather by a diverse set of interdependent dimensions and cultural changes that swept aside the targeted dimension, but in which no single concern was dominant. Hence David Cameron called an up-down referendum on Brexit not in order to re-think Britain's relationship with the EU, but to appease troublesome 'Eurosceptic' MPs who threatened to undermine his

coalition government.

The debate around the referendum, and then the process of implementing the unexpected Leave result, were subjected to powerful, diverse, sometimes contradictory pressures from banks and hedge funds in the City of London, the right-wing press, anti-immigrant activists, powerful political donors, and naked opportunism amongst Cameron's closest friends and allies. British cultural norms shifted from tolerant and cosmopolitan toward nativist and exclusionary. The ultimate result of this complex process was an outcome no one wanted: a 'hard Brexit' that undermined trade with the UK's neighbors and closest trading partners without compensating trade flows with the rest of the world, reduced fiscal space for 'leveling-up' reforms, and substituted migrants from Europe with larger numbers of migrants from countries much further afield.

We should expect the short-term, direct effects of institutional changes shown with the downward arrow in figure 1 to be more predictable than the long-term repercussions of those changes that reverberate through the system. In brief, change begets direct effects and systems effects. Policy makers may predict the former but perhaps not the latter, leading to incongruous outcomes.

We now describe three categories of causes of incongruity that can arise within complex systems: *policy multidimensionality*, *network interactions*, and *interdependent institutions and culture*. Policy multidimensionality consists, in turn, of two sub-categories: *interdependent policy effects*, and *population realignments*. For each category, we present an illustrative model that provides theoretical micro-foundations for each type of incongruity to complement the historical case studies in this SI. We do not claim that this is a complete taxonomy of causes, nor do we see these causes as mutually exclusive. In any real-world case, more than one of these categories of incongruity may be in play, an observation we return to in the paper's final section.

I: Policy Multidimensionality

The most straightforward way that complexity can produce incongruity is through policy multidimensionality. Many (if not most) reforms have effects in more than one dimension. The outcomes and behaviors targeted by a reform will also be affected by what is happening in those other dimensions. So, for example, a reform to decentralize government may bolster the political fortunes of the dictator who decreed it, as happened in Pakistan, or may undermine an established political party system, as happened in Bolivia (Faguet in this SI, Faguet and Shami 2022). The salient point is that different policy dimensions are interdependent in both effects and causes. Traditional analyses often miss such dynamics, whereas complex systems analysis embraces it.

Interdependent Policy Effects

Policy multidimensionality can, in turn, take two forms. The first of these is *interdependent policy effects*, a top-down phenomenon that analyzes how reforms can have effects in multiple dimensions. In these cases, a policy does not miss its target so much as hit other targets too. As above, we assume political actors change institutions with the stated intention of improving outcomes in a *target dimension*. Typical examples of such goals might be reducing crime, increasing economic growth, decreasing corruption, improving educational performance, or improving gender equality. We reiterate that stated goals are used to justify a policy, but may not be leaders' true or main goals.

For simplicity, we assume that absent any other effects, a policy change would be successful. It would result in higher performance by the intended object of reform. In this context, an 'object of reform' is an organization whose behavior / performance is targeted by the reform. For the case of gender quotas in India (Chauchard, Brulé and Heinze in this SI), local governments are the objects of reform; for the case of policing reform in Venezuela (Hanson and Kronick in this SI), objects of reform are teams of police and prosecutors who investigate

and prosecute criminal activity.

For ease of expression, we henceforth refer to objects of reform as *organizations*. We construct a simple model of the following form: Let $x_i(t)$ denote the performance of an organization i on the target dimension at time t. For expository convenience, we assume performance has a minimum of zero and a maximum of 100. Initial performance on the targeted policy dimension is normally distributed with mean 50 and standard deviation σ_V. Performance adapts over time according to the simple updating rule: $x_i(t+1) = x_i(t) + \alpha(x_i(t) - 50) + \epsilon_i$. The parameter $\alpha \in [0, 1]$ captures the rate of adaptation, and ϵ_i is a noise term with mean zero and standard deviation σ_ϵ. If prior to the policy change, α equals zero, then each organization's performance is a random walk.

Borrowing from systems theory, we assume that the policy has two effects: (i) it increases the level of performance, and (ii) it changes the rate of increase in performance. When α becomes positive, the performance updating rule creates positive feedbacks. Organizations with performance above (below) fifty increase (decrease) their performance. Positive feedbacks might be due to learning-by-doing within the organization, and negative ones to growing laziness or shirking, for example. To be effective, the policy change must also shift the performance of every organization above 50. We assume that occurs. Combined with these positive feedbacks, the new policy will cause all organizations to increase their performance.

We now introduce a second outcome dimension influenced by the policy. Abstracting from Chauchard, Brulé and Heinze's (2025) analysis of attempts to increase women's political representation in India, a policy or institutional change can be undermined by changes in a second dimension. For this case, we refer to the targeted dimension as *political equality* and the second dimension as corruption, which we assume the reform increases. To capture this formally, we define *institutional integrity* as the inverse of corruption. We assume that performance in that dimension also takes values between zero and 100, and that prior to the policy change institutional integrity is equal to 50.

Let $y_i(T)$ denote the institutional integrity of organization i, where T is the time of the policy change. Were the policy change to create a *negative externality*, we could write $y_i(T+1) = y_i(T) + X$, where $X < 0$ denotes the size of the externality. Here, we assume the policy change also has a *systems effect* that changes the rate of change in institutional integrity. Formally, for $t \geq T$, we assume that $y_i(t+1) = y_i(t) + \beta(y_i(t) - 50) + \nu_i$, where $\beta > 0$ and ν_i is a normally distributed random variable with mean zero. If X is sufficiently negative that it shifts values of y_i below fifty, then institutional integrity will decrease towards zero for all organizations.

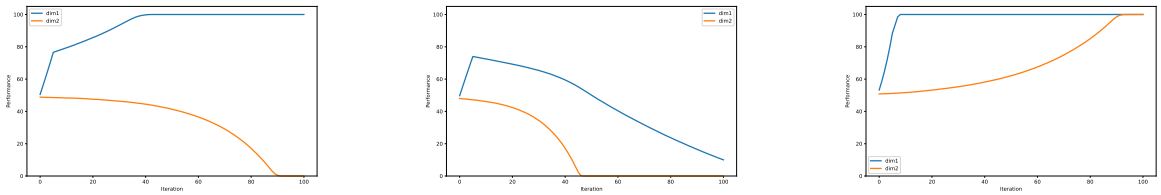
We further assume *systemic interdependence* between policy dimensions. This means that the performance level in one policy dimension depends on performance levels in other dimensions. We might expect that low levels of institutional integrity (high corruption) reduce political equality, and conversely that high levels of political equality increase integrity (Wilkinson and Pickett 2009). To include systemic interdependence in our illustrative model, we can write the two performance updating rules as follows:

$$x_i(t+1) = x_i(t) + \alpha(x_i(t) - 50) + \hat{\alpha}(y_i(t) - 50) + \epsilon_i$$

$$y_i(t+1) = y_i(t) + \beta(y_i(t) - 50) + \hat{\beta}(x_i(t) - 50) + \nu_i$$

where $\hat{\alpha}$ and $\hat{\beta}$ are both positive. The long-term effects of the policy change on the two dimensions now depend on the relative sizes of the positive feedbacks and interdependent effects, as well as on how much the policy change affects levels of equality and integrity. This system of equations can produce a variety of performance patterns. Figure 2 shows three possible dynamic trajectories.

The graph on the left assumes weak interdependence. An example might be the introduction of regional governments in Peru (Eaton in this SI). Here, dimension 1 is Alejandro Toledo's private objective, which was a big reform announcement to help him win a difficult presidential election. In this he was successful, triumphing over two well-known, popular former presidents. But in dimension 2, the reform's stated objective – creating effective regional



(a) Success on Target (b) Failure on Both (c) Success on Both

Figure 2: Interdependent Effects of an Institutional Change or Policy

governments to deepen democracy and better serve Peru's diverse population and geography – the reform was an abject failure.

The middle graph shows a situation similar to Chauchard, Brulé and Heinze's analysis of local governments in India. In many localities, the rise in local female elected officials is more than countered by distortions built into the reform that allow elite males to retain power while using the presence of women to block the empowerment of scheduled castes and tribes. Here, policy interdependence leads to failure in both dimensions.

The graph on the right shows the best-case scenario, which entails interdependence but no incongruity. An example is the introduction of autonomous regional governments in Bolivia (Eaton in this SI). Here, dimension 1 was the stated goal of creating effective, representative regional governments to better serve a very diverse country. Dimension 2 was to reduce political pressures and avert a civil war. The reform was successful on both counts, building a more responsive and accountable state that gained legitimacy in the eyes of citizens, and so drained violence and poison from the central political contest in that country. Success in one dimension was directly connected to success in the second. Or, in more technical language, policy interdependence between the two domains, plus positive feedbacks, led to performance increases in both dimensions.

We can easily expand this model to include more interdependent dimensions responding to a policy change. Assume five dimensions, with a slight downward trend in the four interdependent dimensions and a positive trend in the targeted dimension, and assume that

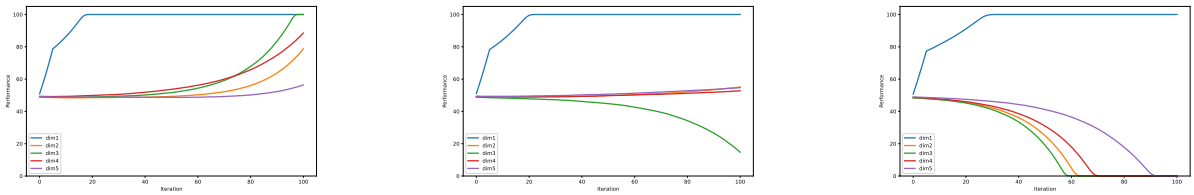
interdependence produces a force towards a common performance level in all five dimensions. One obvious outcome would be that the policy has no effect. Any initial success gets wiped out by interdependent effects, which bring performance in the targeted dimension back in line with other dimensions.

Policy and institutional changes becoming ineffective is a particular, characteristic type of incongruity. In the systems dynamics literature, such *balancing effects* are a major reason why initial policy successes often dissipate (List 2022, Meadows 2008, Sterman 2006). In brief, balancing effects cause outcome variables to return to a natural level. Power differences are a case of particular interest. Groups tend to resist giving up power. Institutional efforts to equalize or reallocate power may be balanced out by system responses. Paniagua and Ricart-Huguet describe how increasing rural regions' electoral power in Argentina was balanced by more cabinet positions for Buenos Aires.

Institutional change will only be effective if the impact of a particular reform is sufficient to induce positive change in other dimensions. This case is shown in the left-hand graph of figure 3. Notice that performance in dimension three crosses performance in dimension four. This occurs because this particular simulation assumes there exists variation in the size of interdependent effects across dimensions. Performance in dimension three was more influenced by performance in the other dimensions.

With more dimensions come more opportunities for incongruity. Consider figure 3(b). Performance in the targeted dimension improves. Three other dimensions also show modest gains. But performance in one dimension falls markedly. We might consider this partial or incidental incongruity. Or it is possible that this was intended by the reforming politician. Also, as shown in figure 3(c), the institutional change could lead to performance declines in all non-targeted dimensions. This would be an extreme example of a policy being effective in its target dimension but leading to worse outcomes overall.

This simple model demonstrates two ways in which cross-dimensional interdependence



(a) Success (b) Partial Success (c) Failure

Figure 3: Multidimensional Interdependent Effects

can produce incongruity. First, systemic interdependence may reverse the intended impact of an institutional change. For example, Venezuela's police reform initially succeeded in greatly reducing wrongful detentions and protecting the rights of suspects and innocents. But it also led to a more than five-fold increase in extrajudicial police killings of suspects they could no longer arrest to investigate. Hanson and Kronick (in this SI) show that this was due to a lack of complementary actions required for reform to succeed. Key failures include no effort to educate the police in forensic methods, nor to hire and train public prosecutors, nor train judges in the new criminal justice system. Success in this interdependent dimension was required for Venezuela's police reform to succeed. But reformers abandoned the judicial side, and so policemen killed suspects instead.

Second, improvement in a target dimension can worsen performance in interdependent dimensions that leave a society worse off overall. Argentina's dual malapportionment is a case in point. Paniagua and Ricart-Huguet argue that the initial legislative malapportionment helped hold a diverse group of provinces with very different interests together. But it also spurred a countervailing executive malapportionment that favored the hegemon, Buenos Aires. Together, the two malapportionments fed a fiscal dynamic in which small, poor provinces hold national governments to ransom in Congress and extract large transfers from larger, more productive provinces. In good economic times, this mechanism is troublesome but workable. But in downturns, it can become pathological and then explosive, as successive Argentine crises have demonstrated.

Population Realignments

Our second type of policy multidimensionality captures situations in which an institutional or policy change creates population realignments. This is a bottom-up mechanism in which a self-organized set of individuals – which might include all the adults or voters in a nation – re-sort themselves into a different pattern when an additional dimension becomes salient. Explaining this type of incongruity requires a quick primer on the key components of *complex adaptive systems*. A complex adaptive system consists of diverse agents whose adaptive behaviors produce micro-level dynamics that in turn produce meso- and macro-level structures and patterns. The patter are referred to as 'emergent phenomena'. These structures and patterns, in turn, influence behavioral responses. Hence, macro phenomena emerge from individual actions and subsequently influence those actions.[1]

The link to incongruity should be clear. An institutional or policy reform can alter emergent structures and patterns. New patterns and structures alter incentives, information, and affordances. These can produce a second category of incongruity if the resulting incentives and behaviors are at odds with a reformed institution's core functions.

We can illustrate how this could occur by constructing an agent based model of two-party spatial (Downsian) electoral competition (Kollman, Miller, and Page 1992). In a spatial electoral model, voters have ideal points represented as real numbers on a set of issues. Candidates from each of two parties take positions on those issues. Voters then vote for the closer candidate in the space. In a model with a three dimensional issue space, a voter with ideal point (4,3,1) would prefer a party proposing the policy (4,2,2) to a candidate proposing (1,6,6) becuase the first candidate's position is nearer to their ideal point.

To build this particular model, we assume four issues. Two are national and two are

[1] Canonical examples include models of racial and income segregation (Schelling 1976), cultural formation (Axelrod 1997), and city-size distributions (West 2018). These patterns are referred to as *self-organized* or *emergent* as they arise without planning. 'Self-organization' and 'emergent' are both used to refer to patterns and structures that arise from the bottom up. Emergence also applies to functionalities that arise at the macro-level, such as consciousness in the brain.

local. We create voter ideal points randomly on each issue assuming a uniform distribution. To initialize the model, we position the incumbent party near the center of the two national issues and farther from the center on local issues. We do the opposite for the challenging party: it advocates more extreme positions on national issues, but more moderate positions on local issues.

Initially, voters only consider national issues. Voters support whichever party is closer in the two national dimensions.Thus, the incumbent party wins handily.

As in a standard agent-based model, we assume that parties maneuver in the space in search of votes. They do so by testing positions near their current position. When a candidate tests a new position, we do not assume that all voters who now prefer that position switch their votes. Instead, we assume that agents switch their vote to the other party if it is closer to their ideal point with a 10% probability. We do so to account for linkages between identity and party affiliation. Note that a rational choice model would assume immediate switching.

If a new position obtains more votes, the party moves its position.[2] Given our assumptions, there is no incentive to move on the local issues. The result of the model will be that the challenger party moves toward the center of the space.

We then alter this standard model to include adaptive behavior by voters and parties. First, we assume that after each election voters move their ideal points on national issues in the direction of their preferred party's ideal point, with the possibility of small deviations.[3] Second, we assume that after each election, parties move their position 2% of theway toward the mean position of their members on that issue. For the first 50 periods, these adjustments only take place on the national issues.

These assumptions produce distributions of ideal points in national and local issues similar to those shown in figure 4, which depicts the agens' positions in the model after another 50

[2]This is called a hillclimbing algorithm.

[3]In the results shown a voter's ideal position on an issue equals 0.98 times its previous position plus 0.02 times its party's position, put a normally distributed error term with mean 0 and standard deviation 0.25

election.

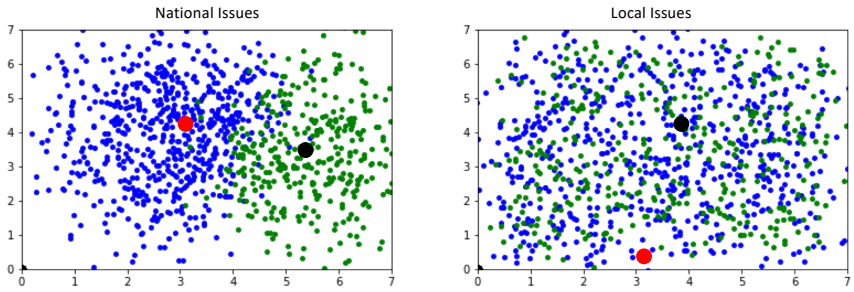


Figure 4: Self-Organization of Ideal Points on National and Local Issues

Look first at the scatterplot on the left. Supporters of party 1 (the incumbent) are colored blue and supporters of party 2 are colored green. The red dot represents the ideal point of party 1; the black dot represents the ideal point of party 2. Two features of this emergent order stand out. First, parties have found their way towards the center of their supporters. Electoral politics produce moderate outcomes. Second, nearly all of the agents support the party that is closer in the national policy dimensions. Given the slowness of the dynamic adjustment process and random errors, the distributions do not converge to all individuals having an ideal point identical to one of the parties. The left side of the figure represents what is called *organized complexity*.

Next, look at the figure on the right showing voters ideal points on the local dimensions. There is no *self-organization* or *emergence*. This represents *disorganized complexity* or randomness (Weaver 1948). There is no correlation between the party that a voter supports and that agent's ideal point.

We now use this model to show how incongruity can arise from realignments. Suppose that party 1 is in power and devolves authority on local issues to local governments (per

Eaton's article in this SI, or Faguet and Shami 2020). Parties and agents will behave in the local dimensions, newly in play, just as they had in the national dimension. They vote for and move toward the closer party. To avoid over-complicating the model, we restrict agents to supporting the same party nationally and locally.

The institutional change has enormous, perhaps unanticipated effects. At first, a modest proportion of voters have an incentive to switch parties. Given our assumptions not all will. Over time, the voters' ideal points and parties' positions in local dimensions also self-organize just as they had in the national dimensions. This is shown in figure 5. The movements by voters and parties in in ideological space and and the new party affiliations change the midpoints of party supporters.

The result is a major realignment. This can be see in figure 5, circles represent parties' ideal points prior to the addition of local issues, and colored discs represent their new positions. Following the change, both parties adopt more moderate positions on local issues.

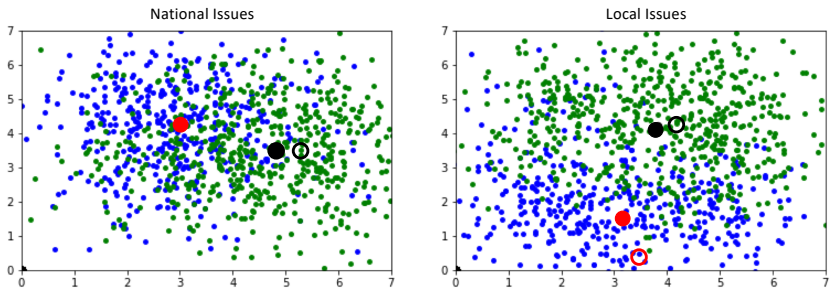


Figure 5: Realignment Following Inclusion of Local Issues

Recall that party 2 began with a slight advantage on local issues. As the adaptive process unfolds, party 2 gains more support than it loses. Since parties move towards their average supporter, new supporters cause party 2 to move closer to the median on national issues.

This results in even more votes. The effects may look subtle in figure 5, but the electoral implications are stark. Figure 6 shows each party's share of the vote before and after activation of the local policy dimension. Party 1's policy change backfires. Party 2 gradually attracts party 1's supporters and eventually achieves a stable majority.

Bolivian politics displays this dynamic but in more extreme form. After the introduction of local governments in 1994, the ruling MNR lost votes and became marginal as voters switched to hyper-local parties and issues. It was Evo Morales' opposition MAS movement that was able to capitalize on the change. The MAS (and its precursors) went from less than 1 percent of the vote in 1997 to between 50-64 percent after 2005 (Faguet 2019).

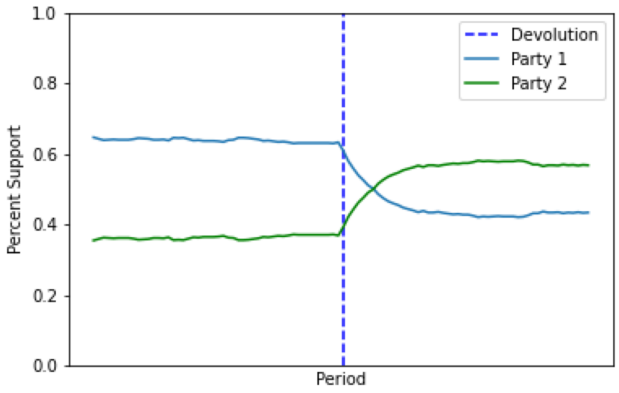


Figure 6: Inclusion of Local Issues Leads to Majority Party Reversal

Ironically, incongruity occurs via self-re-organization. Complex adaptive systems tend to self-organize. Introducing new degrees of freedom, in this case by activating local issues via devolution, results in the system re-organizing from the ground up on account of the new dimension. As Eaton describes in this SI, the creation of elected governorships in Colombia and Peru with the stated goal of strengthening governance unexpectedly broke politics in both countries – in different ways and with different consequences, but in both cases via population realignments. In Colombia, the introduction of gubernatorial elections was intended to open

the nation's politics to non-traditional actors in a bid to transform violent conflict into electoral competition. But the way in which it was done, joining powerful elements of fiscal devolution to political decentralization, undermined national leaders' ability to co-opt emerging leaders and manage the system from above. It also weakened intergovernmental coordination.[4]

The end result was the emergence of new political parties and movements from the regions, which captured local voters and destroyed and replaced a Liberal-Conservative duopoly that had dominated politics and governed the country for 150 years. In Peru, decentralization was half-implemented and then abandoned once Pres. Toledo had won his election. New rules allowed regional movements to contest sub-national elections. This created deep political and fiscal chasms between national parties, which continue to compete in national elections, and the new parties winning regional elections. In Eaton's view, the hierarchical dynamic that has resulted is slowly making Peru ungovernable.[5]

II: Network Interactions

Incongruity can also arise in pockets because many interactions take place over networks (Jackson 2014). If individuals are more likely to interact with people who have similar capabilities, beliefs, and preferences, then clusters of individuals may mutually reinforce behaviors that do not align with policy or institutional objectives. The potential for network and geographic effects to produce clusters of similar behaviors has been the subject of many theoretical and empirical analyses.[6]

To show how local incongruity can arise through network effects, we construct another ABM model of agents who interact in groups of three. Agents choose one of two behaviors: they can be *honest*, h, or *corrupt*, c. The payoff to being honest equals one. The payoff to

[4]For a summary overview of five decades of research on how best to structure central-local fiscal relations, see Faguet and Pal (2023).

[5]A similar, but much more violent, process arguably unfolded in Ethiopia after its local government reforms of the 1990s (Khan et al. 2014 and 2017). This question is ripe for further research.

[6]See Young and Burke (2001) for how agricultural policy can lead to geographic differences.

being corrupt depends on whether or not the agent is caught. We assume an agent is caught if at least one of the other two agents in that group is honest. If caught being corrupt, an agent earns nothing. If not caught, the agent earns two.

To create a baseline, we place agents in random teams of three in each period. We let p be the probability that an agent believes that another agent is corrupt. The payoff to being corrupt strictly exceeds the payoff to being honest if and only if $2p^2 > 1$, or $p > \frac{1}{\sqrt{2}}$. An agent will choose to be honest unless she believes more than 70 percent of other agents to be corrupt. Thus if most agents believe most other agents are not corrupt, all agents behave honestly.

Now we change the model and assume that agents interact not with potentially any other agent in the population, but in small stable groups – i.e. in networks. If the agents with whom a particular agent interacts are mostly corrupt, then that agent will earn a higher payoff by being corrupt.

The phenomenon of different behaviors in different parts of a network is much studied and well understood. To link this to policy incongruity, imagine an institutional reform that reduces previously high payoffs to corrupt behavior to those just described. The expected outcome of such a change would be honest behavior. Corruption would only occur if people believed that a majority of those with whom they interact are corrupt. At the population level that is unlikely, but within parts of a network it could happen.

Magaloni describes such a case in this SI. Criminal justice reforms in Mexico led to clusters of older police officers, who lacked skills in modern techniques of criminal investigation, behaving illegally in ways that directly undermined the stated goals of the reform. Unable to follow new investigative guidelines, they planted evidence to obtain convictions. An institutional change that was expected to improve outcomes produced the opposite through local interactions. But these behaviors did not spread symmetrically across Mexican police forces. They were concentrated in states with low capacity. In high-capacity states, by contrast,

police abuses decreased dramatically and protection of human rights grew stronger.

Another, more dramatic example of network interactions in this SI is Brexit. Hopkin explains how, in 2010, parliamentary support for leaving the European Union was weak. Only a minority of Tories advocated for it, with most Tories and large majorities of Labour, Liberal Democrat, Scottish Nationalist, and other parties' MPs firmly opposed. But facing a hung parliament, and held hostage by Eurosceptic Tory MPs, Prime Minister Cameron gave in to demands for an in-out referendum on leaving the EU. The referendum split his party and then the nation as diverse political, economic, and social actors joined the Leave movement for unrelated reasons. The end result was not just Brexit, but a hard Brexit that left the UK poorer, weaker and more isolated. This case illustrates how isolated pockets of incongruity can infect an entire network. Once part of a network has locked in bad behavior, depending on the characteristics of the network and on social norms, that behavior may spread throughout the network.

A complex systems approach produces two insights. First, it shows how moving from a representative agent perspective – the modal agent is honest, so policy will be effective – to a heterogeneous agent perspective with sorting – some agents will be corrupt, and corrupt agents may be more likely to interact with other corrupt agents – creates the possibility of local incongruity. Second, it shows how local incongruity can spread through a network, leading to general incongruity.

III: Interdependent Institutions and Culture

Incongruity can also arise when a change in an institution or policy produces a cultural change that, in turn, leads to incongruity. We explain how this can occur in two steps. First, we show how cultural change can produce incongruity. By that we mean that a change in culture can produce policy outcomes that were not intended. Second, we show how an institutional change in one policy dimension can affect outcomes in other dimensions via culture. This

differs from interdependent policy effects, which are direct.

We first clarify what we mean by culture. Definitions vary by discipline. Economists typically define culture as values, beliefs, and behaviors (La Porta et al. 1999). Sociologists, political scientists, and organizational theorists include norms and artifacts (Fukuyama 1995, Putnam 1993). Anthropologists add communication structures, symbols, and rituals (Swidler 1986). We conceptualize culture in the way most relevant to institutional performance, as beliefs, behaviors, and norms.[7] We then analyze how culture adapts in response to an institutional or policy change to produce incongruity.

To capture the interplay between beliefs, behaviors, and norms, we modify the *Risk Game* to include norms as shown in figure **??**. The payoff matrix includes a parameter $\Theta_R \in [-1, 1]$ capturing the extent to which norms incentivize risky behavior. The sign of Θ_R determines whether a norm promotes risky or safe behavior. Θ_R negative (positive) implies a larger benefit from taking a safe (risky) action. The absolute value of Θ_R corresponds to the strength of the norm.

	safe	risky
safe	S,S	S,Θ_R
risky	Θ_R,S	$B+\Theta_R$,$B+\Theta_R$

Figure 7: The Risk Game with Norms ($-1 < \Theta_R < 1$)

This construction is consistent with a norm implying a probability of being punished when deviating from expected behavior, be it safe or risky. The norm has a different effect on the payoff structure than changing the payoffs for the risky action because the norm introduces a positive or negative payoff for choosing risky when the other individual chooses safe.

This game connects beliefs, behaviors, and norms. Imagine an individual who holds a belief, p, that others will choose the risky action. For this individual, taking the risk will

[7] By norms we mean the informal rules or behaviors that are socially enforced. This last feature, the inclusion of sanctions, makes them different from beliefs or expectations. Changes to norms can alter payoffs (Bicchierri 2006). The model we describe expands on the *quasi-parameter* construction of Grief and Laitin (2004) in which cultural changes shift payoffs.

produce a higher expected payoff when $p(B + \Theta_R) > S$. A rational actor will thus take the risky action if and only if $p \geq \frac{S}{B+\Theta_R}$. Behavior, both safe and risky, therefore depends on beliefs (p) and norms (Θ_R). If a norm incentivizes risky behavior, ($\Theta_R > 0$), the belief threshold for taking risky action falls. Norms that support risky behavior produce more risky behavior because choosing risk is optimal for a larger range of beliefs. Alternatively, a norm that discourages risk ($\Theta_R < 0$), raises the belief threshold and leads to less risk taking.

In this model, increasing the strength of a norm has a linear effect on the belief threshold, as shown in figure 8. The upward sloping line shows the belief threshold, p_{ra}, for a risk-averse norm as a function of norm strength. The downward sloping line shows the belief threshold for a risk-loving norm, p_{rl}.[8]

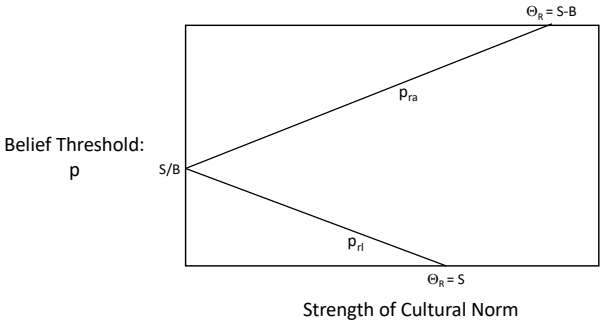


Figure 8: Effect of Cultural Norms on the Belief Threshold

In the two-by-two game setting, the logic is straightforward: whichever behavior the norm encourages becomes more likely. Incongruity could still occur. An institutional change could produce cultural changes that promote less risk taking. As a result, in this game, behavior might shift from risky to safe. Similarly, an institutional or policy change that creates self-interested behaviors, such as allowing people to hold government-subsidized individual retirement accounts, could weaken norms of collectivism in other domains.

While incongruity can occur in two-by-two games, it becomes much more likely in games

[8]Other models might show curvilinear effects of norms on beliefs.

with more actions. To show why this is the case, we analyze a game with *corrupt*, *safe*, and *risky* actions that we call the *CSR Game*. Each of the three actions corresponds to how someone might act in an institutional setting. They might choose the safe action. They might take the risky action that promises the best payoff. This risky action might be to trust other people. Lastly, an individual might deviate and choose the corrupt action. Someone choosing the safe action is insulated from corruption, but someone choosing the risky action is not.

	corrupt	safe	risky
corrupt	1,1	0,0	D,-D
safe	0,0	S,S	S,0
risky	-D,D	0,S	B,B

Figure 9: CSR Game

We first solve for the game's equilibria. The CSR game has three stable pure strategy equilibria. In what follows, we denote the risky strategy b because it produces the best payoff. Basins of attraction in belief space for each of these equilibria are shown in figure 10.[9]

A straightforward exercise shows how these basins of attraction depend on the game's payoff structure. For example, increasing S, the payoff to the safe action, increases the basin of attraction for the equilibrium in which everyone plays safe.

Graphical representations of basins of attraction clarify how beliefs map to behaviors, but do not tell us how a population of adapting agents who began with heterogeneous beliefs collectively arrive in the same basin. In other words, the analysis of how a single point goes to an equilibrium does not fully explain how a population of hetero not always informative about what happens within a complex adaptive system.

To analyze the dynamic updating of beliefs and behaviors in a heterogeneous population, we will rely on an agent-based model. In the simulations that follow, we assume the payoff

[9]All derivations of basin boundaries are in the appendix.

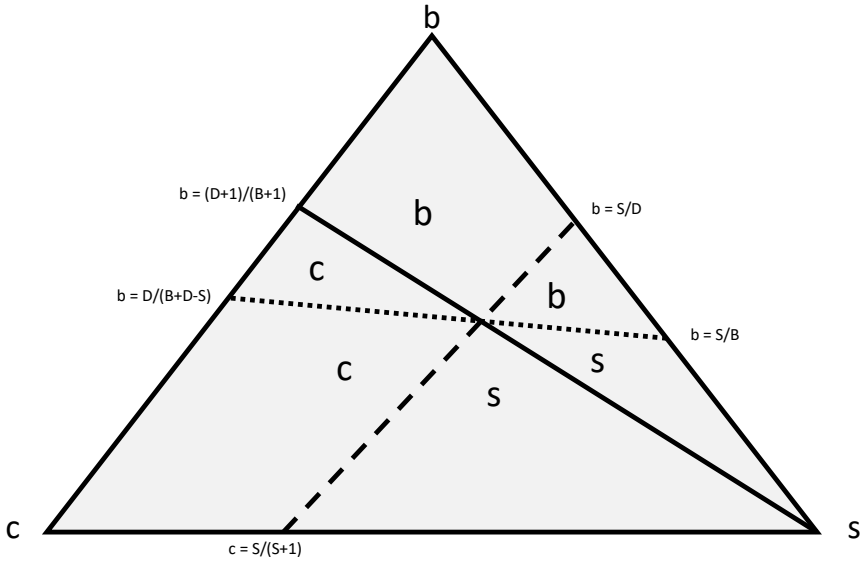


Figure 10: Basins of Attraction for CSR Game

structure shown in figure 11.[10]

	corrupt	safe	risky
corrupt	1,1	0,0	3,-3
safe	0,0	2,2	3,0
risky	-3,3	0,2	7,7

Figure 11: Agent-Based Model Payoffs: CSR Game

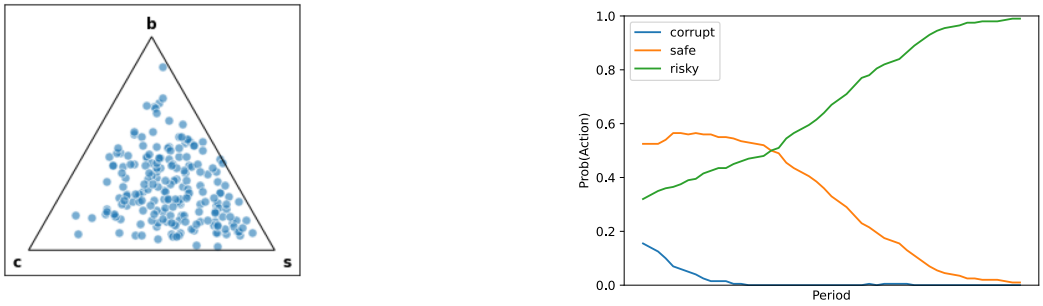
Let $\rho_i^t = (\rho(c_i^t), \rho(s_i^t), \rho(b_i^t))$, denote agent i's beliefs that another player will choose the three actions at time t. By convention, ρ_i^0 equals their *baseline beliefs*.

In the model, assume that agents update their beliefs using the following rule: $\rho_i^{t+1} = 0.85\rho_i^t + 0.05\bar{\rho}^t + 0.10\rho_i^0$, where

$barrho^t$ equals the distribution of mean actions in period t. This equation captures gradual belief updating yet maintains a fixed weight on initial beliefs.

[10]None of the actions is risk-dominant. If all three strategies are played with probability $\frac{1}{3}$, then all have the same expected payoff.

In figure 12a, we show one run of the model with 200 agents, where agents' initial beliefs are biased slightly toward the safe action.[11]



(a) Baseline Beliefs Biased Toward Safe (b) Time Series of Actions

Figure 12: Agent-Based Model With Biased Priors

This initial belief distribution makes the safe action the most likely best response in the first few periods. The decline in the corrupt action along with the rise in the risky action results in even more agents taking the risky action, a classic positive feedback. Eventually, all agents take the risky action.

This single run of the model highlights a crucial difference between agent-based models of dynamical systems and mathematical models. In a mathematical model, we might assume a common initial belief as a proxy for average beliefs in order to make the model tractable. That belief would lie within one of the basins of attraction. The eventual outcome depends only upon this initial belief. If all agents had a belief biased towards the safe action, then all agents would play safe. Referring back to the graph of the basins of attraction, we can imagine an initial belief in the basin of the safe equilibrium that moves towards the lower right-hand corner over time.

In the agent-based model, each agent follows her own belief trajectory. Agents are not playing against a representative agent with mean actions. They are playing against a small number of other agents with diverse beliefs and actions. The outcome therefore depends not

[11]We assume a Dirichlet distribution with parameters (2,3,2).

only on the average belief, but also on the *distribution* of beliefs.

We can employ this model to demonstrate how changes in norms can produce incongruity. Consider first a *risk aversion norm*, which we assume increases payoffs from choosing safe actions and decreases payoffs from choosing risky actions. The change in payoffs results from sanctioning, which is necessary for norms to hold.[12]

	corrupt	safe	risky
corrupt	1,1	$0,\Theta_F$	$D,-D-\Theta_F$
safe	$\Theta_F,0$	$S+\Theta_F,S+\Theta_F$	$S+\Theta_F,-\Theta_F$
risky	$-D-\Theta_F,D$	$-\Theta_F,S+\Theta_F$	$B-\Theta_F,B-\Theta_F$

Figure 13: The CSR Game With Risk-Averse Cultural Norm (Θ_F)

Figure 14 graphs the basins of attraction for this modified game, how these differ from the baseline CSR. Comparing the basins reveals three changes in the basins of attraction caused by the norm change. The basin for the safe action takes region s_b from the risky action basin that gives the best payoff, b. The safe action's basin also adds region s_c from the corrupt action's original basin. Both of these changes might be interpreted as *congruous* in that greater risk aversion results in a greater likelihood of safe actions.

The risk aversion norm also has an unexpected effect: it adds to the basin of attraction for corruption. A region of beliefs, c_b, that had been in the basin of the risky (best) equilibrium is now in the basin of corrupt actions.[13] This occurs even though the payoff to corruption is unchanged. This creates the potential for incongruity. A cultural change that leads to greater risk aversion could increase corruption, depending on the distribution of beliefs. Specifically, if many agents have beliefs in the region c_b, the result would be more corruption.

A similar result could occur with a risk-loving norm. Assume that a risk-loving norm increases the payoff to risky actions and reduces the payoff to safe actions. Figure 15 shows

[12] Assume the risk aversion norm has no effect on the corrupt action.

[13] To apply the formal language of game theory, any agent with beliefs in c_b would previously have best responded by choosing a risky action. With the new norm, her best response will be the corrupt action.

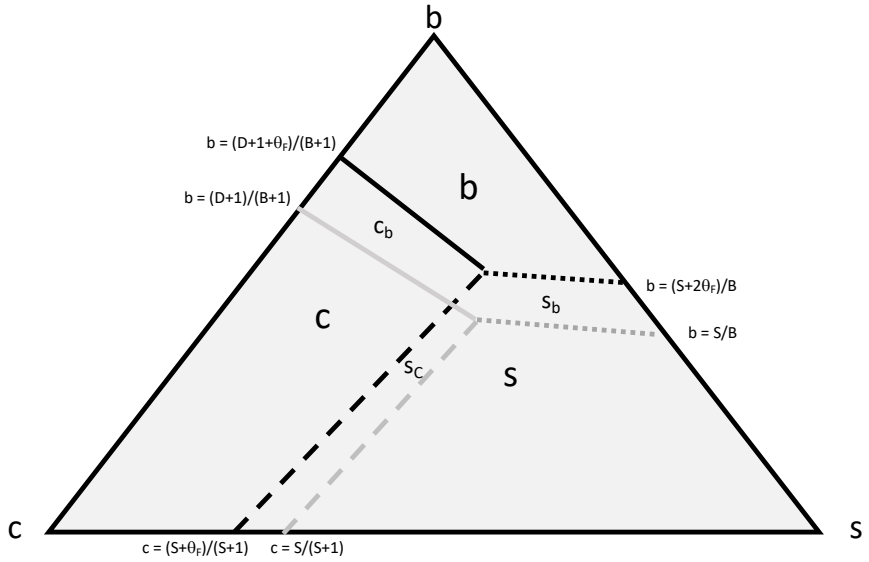


Figure 14: Changes in Basins: Risk Aversion Norm

payoffs in this modification of the game.

	corrupt	safe	risky
corrupt	1,1	$0,-\Theta_R$	$D,-D+\Theta_R$
safe	$-\Theta_R,0$	$S-\Theta_R,S-\Theta_R$	$S-\Theta_R,\Theta_R$
risky	$-D+\Theta_R,D$	$\Theta_R,S-\Theta_R$	$B+\Theta_R,B+\Theta_R$

Figure 15: The CSR Game With Risk-Loving Norm (Θ_T)

Figure 16 shows the basins of attraction for the CSR game when players enforce a risk-loving norm andhow these differ from the baseline CSR game. In this scenario as wel, there exists a region in belief space that produces incongruity. Agents with beliefs in region c_s, will choose corruption rather safety. In sum, risk-avoiding or risk-seeking norms can result – incongrously – in more corruption.

Lastly, we consider a generalized trust norm that punishes corrupt actions when played against risky actions. Figure 18 shows the basins of attraction for a generalized trust norm

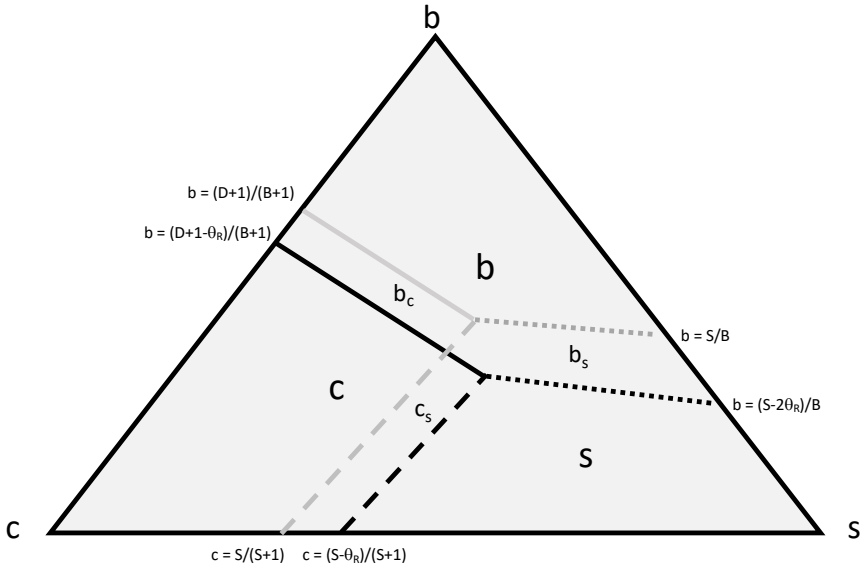


Figure 16: Changes in Basins: Risk Loving Norm

	corrupt	safe	risky
corrupt	1,1	0,0	$D - \Theta_T$,-D
safe	0,0	S,S	S,0
risky	-D,$D - \Theta_T$	0,S	B,B

Figure 17: The CSR Game With Generalized Trust Norm (Θ_T)

and how those differ form the baseline case.

As would be expected, the generalized trust norm reduces the basin of attraction for the corrupt action and increases the basin for the risky action, which produces the highest payoff. But it also increases the basin for the safe equilibrium. We interpret this as a more benign incongruity. An attempt to encourage risk-taking (trusting) behavior by creating norms that punish corruption could result in safe behavior rather than risky (trusting) behavior. Policy makers may not always get what they want (risky behavior), but they might find, sometimes, they get what they need – less corruption.

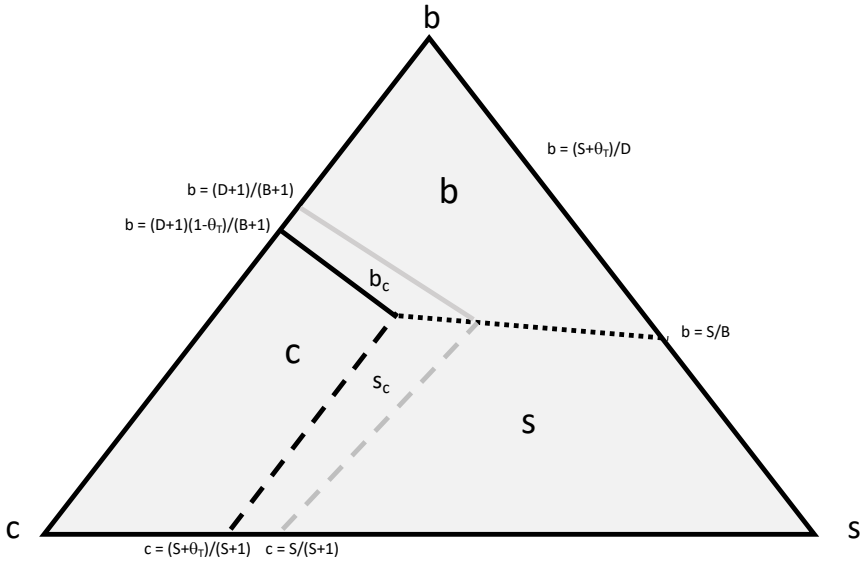
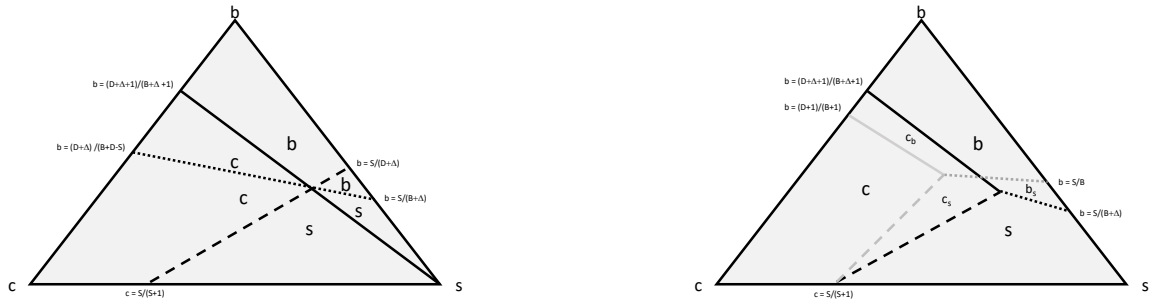


Figure 18: Changes in Basins: Generalized Trust Norm

	corrupt	safe	risky
corrupt	1,1	0,0	$D+\Delta, -D-\Delta$
safe	0,0	S,S	S,0
risky	$-D-\Delta, D+\Delta$	0,S	$B+\Delta, B+\Delta$

Figure 19: CSR Game With Amplified Payoffs

These various thought experiments demonstrate a bottom-up logic for how cultural changes, specifically norms, can produce incongruous outcomes. We now proceed to the second step in our logic by introducing an institutional change that causes a cultural shift that produces an incongruous outcome in another dimension. To do so, we modify our CSR game by adding a parameter, Δ, which we call the *risk amplifier*. Increases in Δ raise the benefit to both players if both choose the risky action, raise gains from corruption if the other player chooses the risky action, and raise the cost of the risky action if the other player chooses corruption. Figure 20a shows the basins of attraction for the amplified CSR game. Figure 20b shows how these differ from the CSR game.



(a) Basins in Amplified CSR Game

(b) Change in Basins from CSR Game

Figure 20: CSR Game With Amplified Payoffs

We can explore the implications of combining generalized trust with amplified risk in the CSR game by building a simple agent-based model featuring a population of agents who play two CSR games, with the payoffs given in figure **??**. Agents have different initial beliefs in the two games, so the paths of beliefs and outcomes differ. We include a generalized trust norm that evolves based on outcomes. We set the initial value of the trust norm, Θ_T^0, to zero. Thereafter, we increase the strength of the norm proportionally to the share of agents who play either the safe or the risky action, and decrease norm strength in proportion to the number of agents choosing the corrupt action. In formal terms, we set $\Theta_T^{t+1} = \alpha\Theta_T^t + (1-\alpha)(\bar{s}^t + \bar{b}^t - \bar{c}^t)$, where $\bar{x}^t$ equals the mean proportion of agents choosing action x at time t across the two games, and $x \in \{s, b, c\}$. Higher values of α imply slower changes to trust norms. In the simulations, we set $\alpha = 0.9$. As Θ_T^t changes, payoffs change per the payoff matrix in figure 17.

We assume initial beliefs in these two games are uniform and allow both actions and the levels of trust to evolve for 200 periods. These parameter settings imply strong incentives for corruption. We then add a third game, an amplified CSR game with $\Delta = 4$. The new game represents an institutional change. We might interpret it as devolving authority to regional governments, reforms to reduce police abuses and improve the administration of justice, or

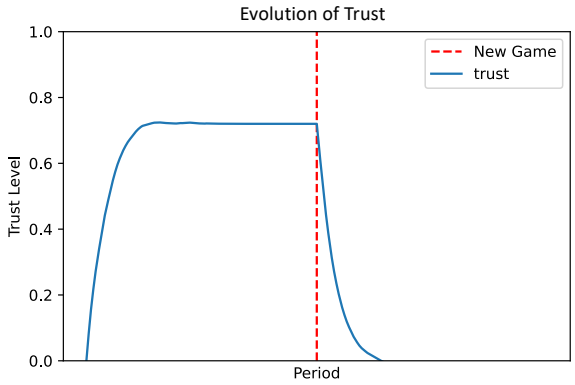


Figure 21: The Collapse of Trust Norm Through Interdependent Effects

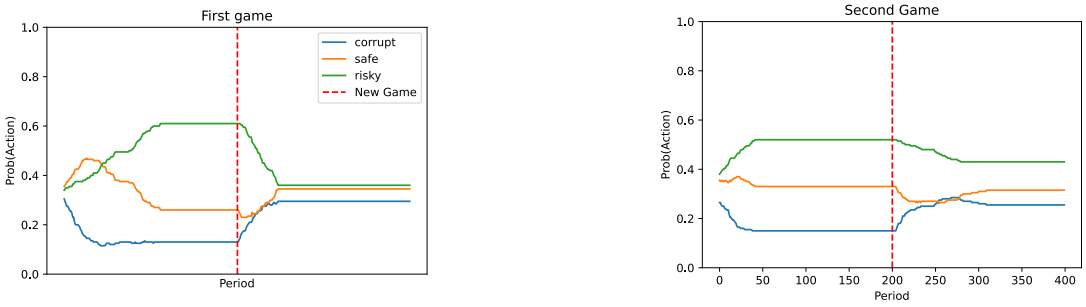
electoral reforms that liberalize party entry. When this new game (institution) is introduced, we assume initial beliefs similar to agents' current beliefs in other games.[14]

In the model, the trust norm applies equally across all games. Formally, this means that Θ_T^{t+1} depends on the mean proportion of agents taking the various actions across all three games. Figure 21 shows the strength of the trust norm over time. The graph shows an increase in trust up to a relatively high level. Then we introduce the amplified risk game (dotted line) and trust drops precipitously.

The collapse of the trust norm produces a system effect by changing payoffs in the other two games. This change in payoffs in turn changes actions through beliefs, as shown in figure 22. Prior to the introduction of the amplified risk game, agents had reached equilibria in the other two games.[15] As evident in the graph, changes in trust occur quickly. Trust collapses. Changes in behavior in the other games then unfold more slowly.

[14] In the results shown, the uniform belief receives a weight of 60% and the average of the other two games is assigned a weight of 40%. See Bednar and Page (2018) for an explanation of how beliefs might spread from one game to another.

[15] In the first game, the equilibrium level of the risky action was higher owing to the distributions of initial beliefs, which were randomly distributed.



(a) Trust Collapse Effects Actions: Game 1 (b) Trust Collapse Effects Actions: Game 2

Figure 22: Institutional-Cultural Interdependence Changing Behavior

In this scenario, culture changed when a new institution was introduced. Culture can also change as a result of the destruction or replacement of an institution. In this SI, Garfias and Sellers describe how the Spanish Empire's expulsion of the Jesuit order from New Spain, on which it had extensively relied to provide colonial education and intermediate with ethnicities who resisted colonization, initially strengthened the Empire's fiscal position. But in the longer term, the expulsion undermined the Crown's legitimacy with many of its native and Creole subjects. This led, in turn, to a decrease in social order, which fed rebellions and insurgencies and ultimately hastened the empire's demise. The effects that Garfias and Sellers describe operated largely through culture. Consistent with our model, the follow-on cultural effects of an abrupt policy change unfolded slowly throughout New Spain, across communities of interacting agents.

3 Discussion: Building Blocks of Incongruity

The purpose of this Special Issue is to document and explore reforms that lead to *institutional incongruity*. Institutions constitute the rules that societies use to govern, allocate resources, make collective decisions, and assign rights and responsibilities (North 1990). If the success

of an institution could be predicted without consideration of place or context using game theory or some other methodology, we would likely see little incongruity. The outcomes of institutional reforms would be known in advance. Instrumental mismatch would become difficult to justify, and incongruous institutions would be scarce. But of course in the world we inhabit, the success of any institution depends to a significant extent on the place and time in which it operates. Specifically, it depends on the ensemble of other institutions in operation, and on local culture. In this SI, Eaton (2025) demonstrates this compellingly in his analysis of direct gubernatorial elections in Colombia, Bolivia, Peru, and Chile. Though similar in principle, each reform played out quite differently because of the other institutions in place in each society, and because each was paired with other institutional and policy changes.

We therefore must not think of simple causal maps of institutional changes that link reforms to outcomes and are common knowledge. Instead, such changes occur within highly complex systems, and the resulting processes are full of uncertainty. Outcomes may be unintentional, or may be intentionally obtuse. Policies may be initially successful, but produce realignments and system effects that lead to worse long-term outcomes. And because of both interdependence and cultural effects, policies that target one goal may have larger impacts on another. Complexity creates fertile ground for incongruity.

Culture holds a special place in our analysis. It is a feature of place that magnifies complexity because it is both shaped by and a shaper of institutions. We believe it necessary to adopt a new approach if we hope to better understand the impact of institutional reforms. We must first widen our analytic lens to consider both ensembles of institutions and culture, and then consider them as interdependent components of a complex system. Each influences the other.

By adopting a complex systems perspective, we can identify multiple pathways through which institutional and policy changes can produce incongruity. This paper analyzes three, one of which has two important sub-categories. We have presented these pathways and pro-

cesses as operating separately. In any real-world setting, of course, incongruity arises through multiple pathways. We thus note, with some irony, that the complex systems approach we have presented oversimplifies the empirical cases in this SI.

Brexit provides an excellent example of multi-causal incongruity. Brexit was promoted by an odd coalition whose members were aiming at different targets. Hopkin describes targets as private goals of policymakers, and others as stated goals. Brexit did reduce financial regulation, a target of the financial sector, and allowed the UK to restrict migrants from Europe. But the consequences for well-being were enormous. Given the scale of the institutional change, it might well be thought a canonical example of *interdependent policy effects*. It had enormous reverberations across the economy. Many of its policy effects fed back into the system, reversing gains in targeted dimensions. In the first three years following Brexit, investment fell by 11% and productivity may have fallen by as much as 5% (Bloom et al 2019). Labor markets tightened; many small businesses that previously exported to Europe went bust; firms that survived spent substantial resources negotiating new trade restrictions rather than on day-to-day management, further contributing to productivity loss; and 67 million British citizens lost free access to live and work throughout Europe. Compared to such losses, the micro gains of improved regulations and more restrictions on some migrants appear small, and a national program to leave the EU looks incongruous.

Brexit also led to a political *realignment* that eventually ejected all the main proponents of Brexit from power. And it had important cultural effects. It raised tensions between British-born and immigrant populations, as well as between the generations. Viewed through a complex systems lens, realignment and cultural change following on the heels of a large institutional change is not surprising.

Another example of multi-causal incongruity is constitutional reform in Chile. Albertus, Menaldo and Rojas-Vallejos describe a complicated, drawn-out set of institutional and political changes that combine cultural effects with population realignment. The setting was

a restrictive constitution and electoral system inherited from the Pinochet dictatorship and intentionally designed to over-represent the political right and hobble the ambitions of the center-left. The 25 years following Chile's return to democracy saw strong growth and development, strong governments, and stable, consensual politics. The feeling grew even on the right that the country had outgrown Pinochet's strictures. Constitutional reforms in 2015 swept aside the old electoral system in favor of proportional representation with public funding for parties. The incentives in the new rules favored smaller, ideologically purer and often more extreme parties, at the expense of the larger heterogeneous political alliances that previously governed the country. These changes in rules and incentives, which led to different electoral outcomes, in turn bled into the political culture. Over time, Chilean politics became less consensual and more fractious as parties became more opportunistic in their behavior.

Two enormous, unexpected shocks then intervened. In 2019, Chilean society exploded in a massive outpouring of frustration with unemployment, inequality, the cost of living, government corruption, privatized pensions, and other issues that crossed the country's political divides. Protesters filled the streets, brought much of the country to a standstill, and made sweeping demands. Then Covid struck. Albertus, Menaldo and Rojas-Vallejos argue that these dual shocks revealed to conservative politicians seeking 'no change' that, in political terms, the population was not where they had thought it was. At the aggregate level, this realization may also have been news to voters. Reforms previously considered fringe now appeared mainstream. And so centrist and right-wing politicians decided to 'join the bandwagon' and vote through constitutional reforms allowing Chileans to access their pension funds during emergencies. Millions of Chileans did. The three massive withdrawals of pension assets that followed totaled 19 percent of GDP, and saw one-third of Chileans fully deplete their pension accounts. How can we understand a conservative government engaging in sweeping reforms that end up wrecking one of the star inheritances from the dictatorship? As the unexpected, incongruous result of population realignments combined with interdependent institutions and

culture.

The potential for cascading incongruity across the political economy deepens the concern raised by Bidegain and Carozzi in this SI that incongruous reforms often target dimensions orthogonal to the general interest. In the case of Uruguay's 1996 electoral reform, the goal was to retain power in the hands of Uruguay's two traditional parties. Bidegain and Carozzi see no reason to presume a general interest here, and we agree. We further see no reason to assume that if such a reform proves incongruous because of a realignment and/or cultural change, the overall effect will be positive. We might instead expect that those with power and resources will benefit. The fact that a reform is instrumental implies that any general interest is incidental to a private goal. Perhaps more importantly, the fact that it is incongruous means that it is also not primarily designed to achieve its stated goals.

Of course, incongruity can also arise from shortsightedness. Many political leaders have opted for short-term gains with little regard for long-term effects. Our claim is that incongruity is far more common than shortsightedness alone would imply. It is especially likely in modern, complex political economic systems because politicians design it in through instrumental mismatch, and because policy multidimensionality and network and cultural effects intervene to make it happen. In a complex system, even well-intentioned policies can become incongruous.

References

Arthur, B. W. 2019. "Complexity Economics: Why Does Economics Need This Different Approach? Complexity Economics" *Proceedings of the Santa Fe Institute's 2019 Fall Symposium. Santa Fe, USA* Santa Fe Institute Press, Santa Fe, NM.

Axelrod, Robert 1997. *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration* Princeton University Press, Princeton, NJ.

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

Bednar, Jenna. 2016. "Robust Institutional Design: What Makes Some Institutions More Adaptable and Resilient to Changes in the their Environment than Others?" In Complexity and Evolution: A New Synthesis for Economics, ed. D.S. Wilson and A. Kirman. Strungmann Forum Reports, vol. 19, J. Lupp, editor. Cambridge, MA: MIT Press.

Bednar, Jenna. 2018. "Modeling the Institutional Matrix: Norms, Culture, and Robust Design." in A Research Agenda for New Institutional Economics, Claude Ménard and Mary M. Shirley, eds., Edward Elgar Publishers.

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.

Bednar, Jenna, Andrea Jones-Rooy, and Scott E. Page. 2015. "Choosing a Future based on the Past: Institutions, Behavior, and Path Dependence."

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. 2016. "Complex Adaptive Systems and Comparative Politics: Modeling the Interaction Between Institutions and Culture." Chinese Political Science Review 1:448-471.

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 and Scott E. Page. 2025. "Institutions and Cultural Capacity: A Systems Perspective." *Journal of Economic Behavior and Organization*, 234:106990,

Bidegain, German and Felipe Carozzi. 2025. "Virtuous outcomes of instrumental intentions? The case of Uruguay's 1996 Constitutional reform."

Boone, P. and J.P. Faguet. 1998. "Multilateral Aid, Politics and Poverty: Past Failures and Future Challenges", chapter 2 in R. Grant and J. Nijman (eds.). *The Global Crisis in Foreign*

Aid, Syracuse: Syracuse University Press.

Bloom, Nicholas, Philip Bunn, Scarlet Chen, Paul Mizen, Pawel Smietanka, and Gregory Thwaites. 2019. "The impact of Brexit on UK firms." National Bureau of Economic Research, No. w26218.

Bowles, Samuel. 1998. "Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions." *Journal of Economic Literature* 36 (1): 75–111

Chauchard, Simon, Rachel Brulé, and Alyssa Heinze. 2025. "Inclusive Reforms as Levers for Social Exclusion: The Paradoxical Consequences of Quotas for Women in Rural India".

Currie, Thomas, Peter Turchin, Jenna Bednar, Peter J. Richerson, Georg Schwesinger, Sven Steinmo, Romain Wacziarg, and John Wallis. 2016. "Evolution of Institutions and Organizations," In Complexity and Evolution: A New Synthesis for Economics, ed. D.S. Wilson and A. Kirman. Strungmann Forum Reports, vol. 19, J. Lupp, editor. Cambridge, MA: MIT Press.

Easterly, William 2017. *The white man's burden: why the west's efforts to aid the rest have done so much ill and so little good.* Old Saybrook, CT: Tantor Media.

Eaton, Kent. 2025. " Electing Governors: The Introduction of Regional Elections in Latin America's Unitary Countries."

Faguet, J.P. 2004. "Building Democracy in Quicksand: Altruism, Empire and the United States." *Challenge: The Magazine of Economic Affairs*, 47: 73-93.

Faguet, J.P. 2019. "Revolution From Below: Cleavage Displacement and the Collapse of Elite Politics in Bolivia." *Politics and Society*, 47 (2): 205–250.

Faguet, J.P. and M. Shami. 2021. "The Incoherence of Institutional Reform: Decentralization as a Structural Solution to Immediate Political Needs." *Studies in Comparative International Development*, 57: 82-109.

Faguet, J.P. and S. Pal (Eds.). 2023. *Decentralized Governance: Crafting Effective Democracies Around the World*. London: LSE Press.

Fisher, R.A. 1930. *The Genetical Theory of Natural Selection.* Oxford, UK: Clarendon Press.

Flyvbjerg Bent and Dan Gardner 2023.*How Big Things Get Done: the Surprising Factors that Determine the Fate of Every Project, from Home Renovations to Space Exploration and Everything in Between.* Pan Macmillan, London, UK.

Fukuyama, Francis. 1995. *Trust*. New York: The Free Press.

Garfias, Francisco and Emily A Sellers. 2025. "The Political Consequences of the Jesuit Expulsion from New Spain."

Gorodnichenko, Y., Roland, G. 2021. "Culture, Institutions and Democratization." *Public Choice* 187, 165–195.

Greif, Avner, and David D. Laitin. 2004. "A Theory of Endogenous Institutional Change" *American Political Science Review* 98 (4): 633–52.

Hopkin, Jonathan. 2025 " The Unintended Revolution: Brexit and Instrumental Incoherence."

Jervis, Robert 1997. *System Effects: Complexity in Political and Social Life.*Princeton University Press, Princeton, NJ.

Khan, Q., J.P. Faguet, C. Gaukler and W. Mekasha. 2014. *Improving Basic Services for the Bottom Forty Percent: Lessons from Ethiopia*. Washington, D.C.: World Bank.

Khan, Q., J.P. Faguet, and A. Ambel. 2017. "Blending Top-Down Federalism with Bottom-Up Engagement to Reduce Inequality in Ethiopia." *World Development*, 96: 326–342.

Kollman, Ken, John Miller, and Scott E Page 1992. 'Adaptive Parties in Spatial Elections." *American Political Science Review*, 86:929–37.

Lamberson, P. J., and Scott E. Page. 2012. "Tipping points.' *Quarterly Journal of Political Science* 7.2: 175-208.

La Porta, R., F. Lopez-de-Silanes, A. Schleifer and R.W. Vishny. 1999. "The quality of government." *Journal of Law, Economics and Organization*, 15: 222-279.

Laitin, David, and Rajesh Ramachandran. 2025. "Switching to English in Rwanda's Educational Curriculum: Estimating the Return on Investment"

Magaloni, Beatriz. 2025. " From Inquisitorial to Adversarial Criminal Justice: Police Brutality and Due Process in Mexico."

Meadows, Donella. 2008 *Thinking in Systems: A Primer*. Chelsea Green Publishing.

Mercier, Hugo, and Dan Sperber, eds. 2017. *The Enigma of Reason.* Harvard University Press, Cambridge, MA.

Miller, John and Scott E Page. 2007. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*, Princeton University Press, Princeton, NJ.

Page, Scott E. 2008. "Uncertainty, Difficulty, and Complexity." *Journal of Theoretical Politics.* 20:115–149. Page, Scott E 2012. "A Complexity Perspective on Institutional Design."

Politics, Philosophy and Economics.11: 5–25.

Putnam, R. D. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton: Princeton University Press.

Sterman, John. 2006. "Learning from Evidence in a Complex World." *American Journal of Public Health* 96, no. 3: 505–515.

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

Weaver, Warren. 1948. "Science and Complexity" *American Scientist* 36(4): 536–44.

West, G. 2018. *Scale: The Universal Laws of Life, Growth, and Death in Organisms, Cities, and Companies*. London: Penguin.

Wilkinson, R. and K. Pickett. 2009. *The Spirit Level: Why Equal Societies Almost Always Do Better*. London: Allen Lane.

Young, H., Peyton, and Mary A. Burke. 2001. "Competition and Custom in Economic Contracts: A Case Study of Illinois Agriculture." *American Economic Review* , 91 (3): 559–573

Proofs

CSR Game

	corrupt	safe	risky
corrupt	1,1	0,0	D,-D
safe	0,0	S,S	S,0
risky	-D,D	0,S	B,B

Figure 23: The CSR Game

The expected payoffs from the three strategies given beliefs (c, s, b) can be written as follows:
corrupt: $\pi_c = c + bD$
safe: $\pi_s = (1 - c)S$
risky: $\pi_b = -cD + bB$

Step 1: $\pi_c = \pi_b$**:** $c + bD = -cD + bB$. It follows that $c = b\frac{B-D}{D+1}$. When $s = 0$, this can be written $(1 - b) = b\frac{B-D}{D+1}$. Rearranging terms gives $(1 - b)(D + 1) = b(B - D)$, which implies $(D + 1) = b(B + 1)$, so that $b = \frac{D+1}{B+1}$. For $s > 0$, payoffs are equal when $c(D + 1) = b(B - D)$. It follows that if $b = 0$, then $c = 0$. Therefore, $\pi_c = \pi_b$ at $s = 1$, as shown by the solid line in figure **??**.

Step 2: $\pi_c = \pi_s$**:** $c + bD = (1 - c)S$. It follows that $c(S + 1) + bD = S$. When $b = 0$, $c = \frac{S}{S+1}$. When $c = 0$, $b = \frac{S}{D}$ as shown with the dashed line in figure **??**.

Step 3: $\pi_s = \pi_b$**:** $(1 - c)S = -cD + bB$. It follows that $S = bB + c(S - D)$. When $c = 0$, $b = \frac{S}{B}$. When $s = 0$, $S = bB + (1 - b)(S - D)$, which gives $b = \frac{D}{(B+D-S)}$ as shown with the dotted line in figure **??**.

To solve for the intersection of the three lines, we substitute $c = b\frac{B-D}{D+1}$ into $c(S + 1) + bD = S$ yielding

$$b\frac{B - D}{D + 1}(S + 1) + bD = S$$

Simplifying gives

$$b[(B - D)(S + 1) + D(D + 1)] = S(D + 1)$$

Rearranging terms gives

$$b = \frac{S(D + 1)}{D^2 + (B - D)S + B}$$

Substituting in $c = b\frac{B-D}{D+1}$, gives

$$c = \frac{S(B - D)}{D^2 + (B - D)S + B}$$

It follows that

$$s = \frac{D^2 - SD + B - S}{D^2 + (B - D)S + B}$$

When $S = 2, D = 3, B = 5$, the unstable intersection $(c, s, r) = (\frac{4}{9}, \frac{2}{9}, \frac{3}{9})$

CSR Game: With Rule Following Norm

	corrupt	safe	risky
corrupt	1,1	$0,\Theta_F$	$D,-D-\Theta_F$
safe	$\Theta_F,0$	$S+\Theta_F,S+\Theta_F$	$S+\Theta_F,-\Theta_F$
risky	$-D-\Theta_F,D$	$-\Theta_F,S+\Theta_F$	$B-\Theta_F,B-\Theta_F$

Figure 24: The CSR Game With Rule Following Cultural Norm (Θ_T)

The expected payoffs from the three strategies given beliefs (c, s, b) can be written as follows:
corrupt: $\pi_c = c + bD$
safe: $\pi_s = (1 - c)S + \Theta_F$
risky: $\pi_b = -cD + bB - \Theta_F$

Step 1: $\pi_c = \pi_b$**:** $c + bD = -cD + bB - \Theta_F$. It follows that $c = b\frac{B-D}{D+1} - \frac{\Theta_F}{D+1}$. When $s = 0$, this can be written $1 - b + \frac{\Theta_F}{D+1} = b\frac{B-D}{D+1}$. Rearranging terms gives $(1 - b)(D + 1) + \Theta_F = b(B - D)$, which implies $(D + 1 + \Theta_F) = b(B + 1)$, so that $b = \frac{(D+1+\Theta_F)}{B+1}$. When $s > 0$ and $c = 0$, payoffs are equal when $bD = bB - \Theta_F$, so $b = \frac{\Theta_F}{(B-D)}$ as shown by the solid line in figure **??**.

Step 2: $\pi_c = \pi_s$**:** $c + bD = (1 - c)S + \Theta_F$. It follows that $c(S + 1) + bD = S + \Theta_F$. When $b = 0$, $c = \frac{S+\Theta_F}{S+1}$. When $c = 0$, $b = \frac{S+\Theta_F}{D}$ as shown with the dashed line in figure **??**.

Step 3: $\pi_s = \pi_b$**:** $(1-c)S+\Theta_F = -cD+bB-\Theta_F$. It follows that $S+2\Theta_F = bB+c(S-D)$. When $c = 0$, $b = \frac{(S+2\Theta_F)}{B}$. When $s = 0$, $S + 2\Theta_F = bB + (1 - b)(S - D)$, which gives $b = \frac{(D+2\Theta_F)}{(B+D-S)}$ as shown with the dotted line in figure **??**.

CSR Game: With Risk Loving Norm

	corrupt	safe	risky
corrupt	1,1	$0,-\Theta_R$	$D,-D+\Theta_R$
safe	$-\Theta_R,0$	$S-\Theta_R,S-\Theta_R$	$S-\Theta_R,\Theta_R$
risky	$-D+\Theta_R,D$	$\Theta_R,S-\Theta_R$	$B+\Theta_R,B+\Theta_R$

Figure 25: The CSR Game With Risk Loving Cultural Norm (Θ_T)

The expected payoffs from the three strategies given beliefs (c, s, b) can be written as follows:
corrupt: $\pi_c = c + bD$
safe: $\pi_s = (1 - c)S - \Theta_R$
risky: $\pi_b = -cD + bB + \Theta_R$

Step 1: $\pi_c = \pi_b$**:** $c + bD = -cD + bB + \Theta_R$. It follows that $c = b\frac{B-D}{D+1} + \frac{\Theta_R}{D+1}$. When $s = 0$, this can be written $1 - b - \frac{\Theta_R}{D+1} = b\frac{B-D}{D+1}$. Rearranging terms gives $(1 - b)(D + 1) - \Theta_R = b(B - D)$, which implies $(D + 1 + \Theta_R) = b(B + 1)$, so that $b = \frac{(D+1-\Theta_R)}{B+1}$. When $s > 0$, and $b = 0$, payoffs are equal when $c = -cD + \Theta_R$, so $c = \frac{\Theta_R}{D+1}$ as shown by the solid line in figure **??**.

Step 2: $\pi_c = \pi_s$**:** $c + bD = (1 - c)S - \Theta_R$. It follows that $c(S + 1) + bD = S - \Theta_R$. When $b = 0$, $c = \frac{S-\Theta_R}{S+1}$. When $c = 0$, $b = \frac{S-\Theta_R}{D}$ as shown with the dashed line in figure **??**.

Step 3: $\pi_s = \pi_b$**:** $(1-c)S-\Theta_R = -cD+bB+\Theta_R$. It follows that $S-2\Theta_R = bB+c(S-D)$. When $c = 0$, $b = \frac{(S-2\Theta_R)}{B}$. When $s = 0$, $S - 2\Theta_R = bB + (1 - b)(S - D)$, which gives $b = \frac{(D-2\Theta_F)}{(B+D-S)}$ as shown with the dotted line in figure **??**.

CSR Game: With Generalized Trust Norm

	corrupt	safe	risky
corrupt	$1 - \Theta_T$,$1 - \Theta_T$	$-\Theta_T$,0	$D - \Theta_T$,-D
safe	0,$-\Theta_T$	S,S	S,0
risky	-D,$D - \Theta_T$	0,S	B,B

Figure 26: The CSR Game With Generalized Trust Norm (Θ_T)

	corrupt	safe	risky
corrupt	1,1	,0	$D - \Theta_T$,-D
safe	0,$-\Theta_T$	S,S	S,0
risky	-D,$D - \Theta_T$	0,S	B,B

Figure 27: The CSR Game With Generalized Trust Norm (Θ_T)

The expected payoffs from the three strategies given beliefs (c, s, b) can be written as follows:
corrupt: $\pi_c = c + bD - b\Theta_T$
safe: $\pi_s = (1 - c)S$
risky: $\pi_b = -cD + bB$

Step 1: $\pi_c = \pi_b$**:** $c + bD - b\Theta_T = -cD + bB$. It follows that $c(D+1) - b\Theta_T = b(B - D)$. When $s = 0$, this can be written $(1-b)(D+1) - b\Theta_T = b(B-D)$. Rearranging terms gives $D + 1 - b\Theta_T = b(B + 1)$, which implies $(D + 1) = b(B + 1 + \Theta_T)$, so that $b = \frac{(D+1}{B+1+\Theta_T}$.

When $s > 0$ and $b = 0$, payoffs are equal when $c = -cD$. Therefore, $c = 0$ as shown in the solid line in figure **??**.

Step 2: $\pi_c = \pi_s$**:** $c + bD - b\Theta_T = (1 - c)S$. It follows that $c(S + 1) + bD = S + b\Theta_T$. When $b = 0$, $c = \frac{S}{S+1}$. bWhen $c = 0$, $b = \frac{S}{D-\Theta_T}$ as shown with the dashed line in figure **??**.

Step 3: $\pi_s = \pi_b$**:** $(1 - c)S = -cD + bB$. It follows that $S = bB + c(S - D)$. When $c = 0$, $b = \frac{S}{B}$. When $s = 0$, $S = bB + (1 - b)(S - D)$, which gives $b = \frac{D}{(B+D-S)}$ as shown with the dotted line in figure **??**.

CSR Market Game:

	corrupt	safe	risky
corrupt	1,1	0,0	$D + \Delta, -D - \Delta$
safe	0,0	S,S	S,0
risky	$-D - \Delta, D + \Delta$	0,S	$B + \Delta, B + \Delta$

The expected payoffs from the three strategies given beliefs (c, s, b) can be written as follows:

corrupt: $\pi_c = c + b(D + \Delta)$

safe: $\pi_s = (1 - c)S$

risky: $\pi_b = -c(D + \Delta) + b(B + \Delta)$

Step 1: $\pi_c = \pi_b$**:** $c + b(D + \Delta) = -c(D + \Delta) + b(B + \Delta)$. It follows that $c = b\frac{B-D)}{D+\Delta+1}$. When $s = 0$, this can be written $(1 - b) = b\frac{B-D}{D+\Delta+1}$. Rearranging terms gives $(1 - b)(D + \Delta + 1) = b(B - D)$, which implies $(D + \Delta + 1) = b(B + \Delta + 1)$, so that $b = \frac{D+\Delta+1}{B+\Delta+1}$. For $s > 0$, payoffs are equal when $(1 - s - b)(D + \Delta + 1) = b(B - D)$. It follows that if $b = 0$, that $s = 1$. Therefore, $\pi_c = \pi_b$ at $s = 1$, as shown by the solid line in figure **??**.

Step 2: $\pi_c = \pi_s$**:** $c + b(D + \Delta) = (1 - c)S$. It follows that $c(S + 1) + b(D + \Delta) = S$. When $b = 0$, $c = \frac{S}{S+1}$. When $c = 0$, $b = \frac{S}{D+\Delta}$ as shown with the dashed line in figure **??**.

Step 3: $\pi_s = \pi_b$**:** $(1-c)S = -c(D+\Delta)+b(B+\Delta)$. It follows that $S = b(B+\Delta)+c(S-D\Delta)$. When $c = 0$, $b = \frac{S}{B}$. When $s = 0$, $S = b(B+\Delta)+(1-b)(S-D-\Delta)$, which gives $b = \frac{D+\Delta}{(B+D-S)}$ as shown with the dotted line in figure **??**.

CSR Market Game With Risk Loving and Generalized Trust Norms

	corrupt	safe	risky
corrupt	$1 - \Theta_T, 1 - \Theta_T$	$-\Theta_T, -\Theta_R$	$D + \Delta 1 - \Theta_T - \Theta_R, -D - \Delta +$
safe	$-\Theta_R, -\Theta_T$	$S - \Theta_R, S - \Theta_R$	$S - \Theta_R, \Theta_R$
risky	$-D - \Delta + \Theta_R, D + \Delta - \Theta_T - \Theta_R$	$\Theta_R, S - \Theta_R$	$B + \Delta + \Theta_R, B + \Delta + \Theta_R$

The expected payoffs from the three strategies given beliefs (c, s, b) can be written as follows:

corrupt: $\pi_c = c + b(D + \Delta) - \Theta_T$

safe: $\pi_s = (1-c)S - \Theta_R$
risky: $\pi_b = -c(D+\Delta) + b(B+\Delta) + \Theta_R$

Step 1: $\pi_c = \pi_b$**:** $c+b(D+\Delta)-\Theta_T = -c(D+\Delta)+b(B+\Delta)+\Theta_R$. Combining like terms give $c(D+\Delta+1) = b(B-D)+\Theta_R+\Theta_T$. When $s=0$, this can be written $(1-b)(D+\Delta+1) = b(B-D)+\Theta_R+\Theta_T$, which simplifies as $(D+\Delta+1) = b(B+\Delta+1)+\Theta_R+\Theta_T$, so that $b = \frac{D+\Delta+1-\Theta_R-\Theta_T}{B+\Delta+1}$. When $b=0$, the two payoffs are equal when $c(D+\Delta+1) = \Theta_R+\Theta_T$, implying that $c = \frac{\Theta_R+\Theta_T}{D+\Delta+1}$ as shown in the solid line in figure **??**

Step 2: $\pi_c = \pi_s$**:** $c+b(D+\Delta)-\Theta_T = (1-c)S-\Theta_R$. It follows that $c(S+1)+b(D+\Delta) = S+\Theta_T-\Theta_R$. When $b=0$, $c = \frac{(S+\Theta_T-\Theta_R)}{S+1}$. When $c=0$, $b = \frac{S+\Theta_T-\Theta_R)}{D+\Delta}$ as shown with the dashed line in figure **??**.

Step 3: $\pi_s = \pi_b$**:** $(1-c)S - \Theta_R = -c(D+\Delta) + b(B+\Delta) + \Theta_R$. It follows that $S - 2\Theta_R = b(B+\Delta) + c(S-D\Delta)$. When $c=0$, $b = \frac{S-2\Theta_R}{(B+\Delta)}$. When $s=0$, $S-2\Theta_R = b(B+\Delta)+(1-b)(S-D-\Delta)$, which gives $b = \frac{D+\Delta-2\Theta_R}{(B+D-S)}$ as shown with the dotted line in figure **??**.