Public Choice
https://doi.org/10.1007/s11127-019-00705-4

Game theory and the study of American political development

Sean Gailmard[1]

Received: 31 July 2019 / Accepted: 6 August 2019

Abstract
Game theoretic analyses of American institutions and American political development largely are disconnected enterprises, yet they share many points of contact and thus opportunities for fruitful exchange. In this essay I discuss the value and limits of formalization for the enterprise of institutional analysis that those fields have in common. I conceptualize two broad approaches that formal modelers have taken to study institutions—institutions as game forms, and institutions as equilibria—that have been relatively successful for understanding institutional *choice* and *stability*. At the same time, formal modelers have been less successful in addressing institutional *change* and *development*, topics about which APD has much to offer. Overall, I contend that crosstalk between the two fields can benefit them both.

Keywords Game theory · American political development · Political economy

JEL Classification C79 · D70

Within political science, few intellectual gulfs would seem larger on first sight than that separating formal political theory and American political development (APD). One of the research traditions chiefly uses a mathematical idiom, emphasizes logical derivation, and prizes simplicity in the representation of political processes. The other primarily is verbal, often based on archival material, and tolerates complexity of representations easily. One field aspires to generality and crisp, sometimes ironic or even "counterintuitive" results; the other emphasizes contingency and accepts particularity when necessary to explain an important development. One field has most of its intellectual crosstalk with the discipline of economics, the other with history and historical sociology.

In view of that yawning divide, it is perhaps unsurprising that very little research exists that straddles the boundary between the two research traditions. From the paucity of collaboration, one might even surmise that productive engagement across the two fields, like the proverbial $10 bill that the economist's child spots on the ground, is out of equilibrium—thus, the search for either one is fruitless and best not attempted.

✉ Sean Gailmard
gailmard@berkeley.edu

[1] University of California, Berkeley, USA

Published online: 09 August 2019

I believe that the foregoing conclusion is wrong and so, in this paper, I will make the opposite argument: that important intellectual gains can be captured from exchange across these fields. In particular, I argue that the potential for a two-way dialogue between formal modeling and APD is substantial. Each can contribute to the other.

My completely subjective prior belief is that both modelers and APDers readily will accept the first contention of that argument, that APD can contribute to formal theory. Formal theorists well understand their dependence on others to furnish questions and puzzles that are interesting and important. For their part, APDers presumably are inclined to believe their field has no shortage of them. APD often presents puzzles and insights around which new models can be built. When those puzzles have not been subject to previous formal analysis, they present an opportunity for original research within formal theory. That is intellectually desirable because representation in a model highlights the portability of a puzzle or idea to other contexts, making it possible to spot other instances of the same puzzle, even in radically different substantive contexts and thereby better grasping the fundamental driving forces of politics.

But at least some colleagues may doubt the second contention, that formal modeling has something useful to contribute to APD. After all, APD is not a body of questions or data, lying inert until given life by application of a suitable and rigorous method. To its practitioners, APD is both a topic and an approach (Orren and Skowronek 2004). It emphasizes the process of change in large-scale institutions or policies, often produced from the abrasions or collisions of multiple distinct actors and institutions. A formal model will not necessarily provide insights that such scholars seek; if so, designing a model around events in the past will not change that state of affairs.

Yet I believe that here, too, beneficial interchange is possible. Formal models can make abstract processes clearer and more concrete, features prized by all scholars in theoretical work. Formal theory also has several productive idioms for expressing the contours of an "institution." Since much of APD focuses on deep processes by which political institutions are established, evolve, or reproduce themselves, some scope exists for formal modeling to contribute to the APD literature. Formal representation of an argument also sometimes helps to expose additional conditions and dependencies beyond those captured in a verbal version of it, representing another channel by which formal modeling can potentially contribute to APD.

In the remainder of this paper, I develop that argument by briefly considering the nature of game theoretic modeling in general. I follow that argument by discussing, in turn, two loosely differentiated approaches to modeling institutions in game theory: institutions as extensive forms, and institutions as equilibria of some more fundamental game. I explore how each approach has been used to develop insights about APD with specific examples from the literature. Following that discussion, I turn to a variety of critiques of game theoretic modeling leveled by APDers and historical institutionalists. I argue that some of those critiques are accurate and important, while others reflect misunderstandings of game theory. Both types of critique present challenges to formal theorists interested in the APD field: the first, a challenge to improve our models; the second, to improve our communication about them. I conclude by summarizing and offering suggestions for future directions.

In making that argument, owing to space constraints I limit attention to research that contains a formally specified, game theoretic model of some form about the development of US institutions.[1] All of the models summarized are grounded in rational choice theories in the sense that actors have transitive preferences over outcomes and the ability to solve the dynamic programming problem embedded in the game (or behave as if they do). I do not include APD research with verbal arguments based on "rational choice theory" or inspired by ideas from formal models developed elsewhere. That broader body of research is somewhat large; for discussion and review, see Jenkins (2016). I also focus specifically on modeling, not on the construction of historical narratives in light of a model's discipline (cf. Bates et al. 1998).

1 Decisions, institutions, and models

To evaluate the potential contributions of formal theory to APD, it is necessary briefly to fix some ideas of exactly what formal theory is about and how its practitioners work. The technicalities are better expounded elsewhere, so I will be mercifully brief about them. I spend more space considering what formal modeling does and does not require, since some confusion may exist about those requirements, and give a sense of how at least one formal theorist thinks about research questions.

Broadly speaking, formal models conceive of institutions in one of two ways: either as the extensive form of a game, or as the outcome of some other, larger, more fundamental game (Shepsle 1989; Calvert 1995). An extensive form defines a set of players' feasible actions in a game, how those actions are sequenced, and what each decision maker or player knows about the potential actions and preferences of other players when they act. Any sequence of actions taken throughout the extensive form, or "path of play," terminates in an outcome. Players have complete and transitive preferences over outcomes. Practically speaking, those preferences are represented by a utility function.

The analysis of games focuses primarily on one question: what is a sensible course of action for each player at each decision node in the game? A variety of ways are available for answering that question, but the answer almost always involves identifying Nash equilibria, or patterns of action in which no player knowingly acts to induce a less preferred outcome than another that is possible, given the action plans of all other players. That equilibrium condition is equivalent to each player acting to maximize her utility from outcomes given their beliefs of what others do, and holding beliefs that are consistent with what others actually do.[2]

Game theoretic modeling excels at explaining situations when some actor or group of actors takes a decision that appears to be puzzling, which often means that it seems detrimental to the decision makers in a way they should understand. One facet of such puzzles is that a radical disconnect may exist between what the decision maker desires, and what she actually obtains in equilibrium—a disconnect evinced, for example, in classic collective action and free-riding games. Or a puzzle may lie in an actor's decision to contribute to

[1] Thus, I do not consider social choice or decision theoretic models, e.g. Miller and Schofield (2003), Schofield (2006), and Ballingrud and Dougherty (2018).

[2] Nash equilibrium thinking already may seem unacceptable when one wishes to argue that an event or institutional change occurred because one actor did something another did not think was possible, or that an outcome occurred that some actor did not realize was possible. The essence of such an explanation is that some actor(s) have incorrect beliefs—either about another's decisions, or about the game itself. I will return to that issue and discuss formal representations of it below.

financing a collective good, when they have obvious incentives not to contribute. Another source, relatively common in formal models of APD, is that one actor cedes power to a second actor who may, through that grant of power, be able to take advantage of the first.

Strategic interaction has a way of flipping the accounts of which of the available actions obviously are beneficial or obviously detrimental to an actor's interests. To a formal political theorist, when a puzzling decision is encountered, the challenge is to identify or invent a strategic context—an extensive form game—in which the seemingly detrimental action actually was beneficial, given the actions of other players. Such strategic considerations resolve the puzzle by rationalizing the decision, for a rational decision is per se not puzzling.[3] "The decision maker took a detrimental action because the decision maker is an idiot" never is a good explanation; saying that, "the decision maker, who lived and breathed the decision problem in its full context, did not understand it as well as I, the analyst" is very close to "the decision maker is an idiot."

The discussion just presented presumes that the modeling effort begins with a real-world puzzle, and attempts to devise an extensive form game ("a model") that resolves it. Criteria for evaluating whether research in that spirit is "good" include whether the puzzle is substantively interesting; whether the puzzle seems to recur in a variety of contexts; whether the model so devised is new to the literature, not simply a trivial change from some other model; and whether the model resolves the puzzle in a strategically evocative (or even, sometimes, "counterintuitive") manner.

To be sure, not all formal theory research proceeds in that way. Some of it is driven by the formal theory literature itself, e.g., changing a particular assumption of a model and showing that the results differ qualitatively (or, if they are not, therefore concluding that the assumption is not doing the heavy lifting). Literature-driven research sometimes may be important to the formal theory field internally, but is not a major source of connection to APD. To build that connection, the most successful approach is to build models around substantive puzzles in APD, and that is the literature I consider below.

1.1 Benefits of formalization

Although related, reasoning game theoretically is different from formally stating and analyzing a game theoretic model. My argument is not just about game theoretic reasoning, but formalization. Formalization can bring two possible benefits: communication of ideas, and generation of ideas. Those benefits arise because formalization requires clarity of assumptions, and forces consistency of assumptions and conclusions.

In terms of communication, consider a theorist who has specified and analyzed a formal model. The model embeds a causal process and specifies the key variables involved; it specifies how those variables interact to determine an outcome of interest. Although the theorist has used mathematical tools to express a causal argument and derive relationships among variables, she does not have to communicate that argument or insights to her colleagues in formal, mathematical terms. She could instead use a natural language description of the variables, the relationships, and the causal process relating them, but leave the formal model in the desk drawer.

[3] A decision maker also may take a detrimental decision that s/he has no reason to know is detrimental. Such action also is not puzzling, but "the decision maker did not know any better" usually does not make for an interesting explanation, unless there is some reason to expect that they should have.

Communicating with a model involves a tradeoff of accessibility and precision. On one hand, anyone unable to parse mathematical expressions in general, or unfamiliar with the language and techniques of game theory, will find the presentation inaccessible. On the other hand, for readers past that barrier to entry, communication in mathematical idiom can enhance the clarity of the ideas being presented. For example, suppose a theory involves a group of disparate individuals acting in a unified way, say "an interest group threatens a protest." When the theory is presented in a formal model, it is impossible to elide the assumption that the individual members of the group have been summarily imbued with a common interest. That, in turn, can bring needed focus to how that common interest comes about. When the theory is presented in verbal description, a skilled writer can more easily gloss over that assumption or make it appear as an ineluctable truth of such groups.

In terms of generation of ideas, the question is whether to execute the analysis in a mathematical form in the first place—bracketing the issue of how the ideas will be communicated. Here, the theorist's choice is how to develop the ideas. She might use a formal, mathematical approach; or might instead opt for a purely verbal mode of analysis. At that stage, two benefits to formalization emerge: it forces the theorist to be clear and concrete about the variables and causal processes she actually is considering, and it forces the analyst to be logically consistent in deriving conclusions from premises.

Everyone values clarity and consistency; the question for every theorist is whether she or he has achieved them. A danger of verbal arguments is that they might look convincing to their creator, but his abiding belief in his own abilities as a theorist may cloud his judgment of his own performance. Or an informal argument might inadvertently assume contradictory motivations for the actors involved. Another possibility is that an informal argument relies on hidden assumptions that the analyst did want to make or realize he was making. Formalization eliminates those problems; if the conclusions do not follow from the stated premises, or the premises do not deliver the conclusion desired, the theorist is forced to confront it when the proofs don't work.

Needless to say, examples abound to prove that formalization is neither necessary nor sufficient for clarity or consistency. My contention is simply that sometimes it helps.

1.2 Limitations

Two important assumptions in equilibrium analysis of games limit the range of theoretical explanations it can express. First, Nash equilibrium play in a game not only prescribes rational action in light of given beliefs about what others will do; it prescribes that the beliefs are correct.[4] To use equilibrium as an analytical tool is to assume belief consistency that is not explained by the model itself.

That may be the most significant assumption one must accept when analyzing games in terms of equilibria. If incorrect beliefs of agents about others are integral to a theory of political action, then standard Nash equilibrium analysis is not well suited for expressing or probing that theory. Other modes of analysis allow incorrect beliefs about what others would do at contingencies that never arise (self-confirming equilibrium), or place very weak consistency restrictions on beliefs (rationalizability). At least one paper in the formal

[4] In Nash equilibrium and perfect Bayesian equilibrium, beliefs must be correct at all information sets; in self confirming equilibrium, beliefs must be correct only at information sets on the equilibrium path of play. Belief consistency does immense work beyond rationality alone to reduce the set of actions it is reasonable for players to take.

APD literature employs the former concept (Defigueiredo et al. 2006), but both are uncommon in applications to date. Nevertheless, it is important to point out that one can execute formal analysis and obtain some of its benefits, such as clarity of assumptions about actors' motivations, without assuming the belief consistency of Nash equilibrium. Sometimes simply writing down utility functions is helpful for clarifying an argument.

Second, any formal analysis based on decision theory is, at present at least, unable to incorporate the concept of an unforeseen contingency. No method has yet been discovered to represent unawareness in standard models of choice under uncertainty.[5] It is simply not clear how to think about how actors think about things they cannot think about, whether one prefers a verbal or a mathematical idiom. So, if unforeseen contingencies are integral to a theoretical explanation, then decision-theoretic tools (and per force game theory) are unsuited for expressing it. Unlike belief consistency, unforeseeable events are not just an issue with Nash equilibrium; they raise issues with expected utility theory.[6]

On the other hand, it is not clear that any verbal mode of theorizing does a better job of handling unforeseen contingencies. It is easy to talk about the concept, and to offer anodyne prescriptions such as, "we should think of institutional designers as making their institutions robust to unforeseen contingencies." What is never quite clear in a verbal account is whether that prescription does important analytical work. In the economic theory of incomplete contracts, wherein unforeseen contingencies are a major issue, precise formalizations suggest that it does not (Maskin and Tirole 1999).[7]

In critiques of game theoretic modeling, it is important to understand that it is not a static field. Objections to one class of techniques or to common modeling approaches should not be taken as objections to the approach in general. First, they may reflect transitory technical limitations in the field, which can be obviated by future developments. For instance, when Downs (1957) wrote *An Economic Theory of Democracy*, rational choice theory barely had developed tools for evaluating choice under uncertainty, and game theory was embryonic. Critiques of Downs's modeling approach as unable to deal with strategic interaction or uncertainty therefore were nullified by technical progress in the field. Second, modeling techniques in use by theorists may reflect the idiosyncratic tastes of people who happen to populate the field at a given time, which evidently can change.

[5] However, a crucial distinction must be made between an unforeseen contingency and a zero-probabability event, the analysis of which is fairly common in games of incomplete information. To see the difference, suppose that a decision maker is authoritatively told "the event *X* might or might not happen." If she was aware of *X*, that statement is uninformative. If she was unaware of *X*, it is informative, and might affect her decisions. That situation is roughly analogous to Treasury Secretary Steven Mnuchin's declaration in December 2018 that "there is no liquidity crisis in American banking." His statement rattled financial markets because investors had not even considered it, but the declaration suggested that they should.

[6] Note, however, that unforeseen contingencies and "unintended consequences" are quite different. The former may sometimes imply the latter, but is not necessary for it. For instance, in the Prisoners' Dilemma game, it seems clear that jointly producing the worst possible collective outcome is not either player's intention. The interesting thing is exactly that it happens despite no one intending it. But equally clearly, that outcome is not unforeseen. Thus, if one's theoretical interest is really in the unintended consequences of institutions, it is important not to conclude that the inability of decision-theoretic models to capture unawareness renders them inapplicable.

[7] Consider the possibility of traffic delays faced by a commuter in the morning rush hour. Unforeseen contingency models might assume that the commuter did not realize it was possible to be delayed 30 minutes by a spill from a molasses truck. They do assume that the commuter knows that it is possible to be delayed by 30 minutes—just not the full list of the reasons why. But if all possible payoffs are foreseeable, it is not clear that the unforeseen contingency of a molasses truck spill is doing any work.

Those technicalities aside, it sometimes is contended that the real limitation of game theory is that it forces assumptions that are particularly inappropriate in the study of political development, or that some types of theories simply cannot be expressed in its language (Pierson 2004). Even illustrious formal theorists have subscribed to that view, e.g., "Game-theoretic accounts require detailed and fine-grained knowledge of the precise features of the political and social environment within which individuals make choices and devise political strategies" (Bates et al. 1998, p. 628). Thus, if one lacks such fine-grained knowledge of precise features, one ostensibly cannot employ a game theoretic account.

Likewise, Pierson and Skocpol (2002, p. 11) contend, "Game theory generally requires that all the relevant actors, preferences, and payoffs be established and fixed simultaneously at the beginning of a game". Thus, the tool is held to be inapplicable when preferences change, new players emerge, or established players preclude the appearance of new ones.

The position of the present author is that such views are mistaken. Specifically, no theory of *intentional* action in politics exists—a theory in an "action frame of reference"—that cannot be cast in game theoretic terms. What that requires is not a list of players and preferences fixed ex ante, or a "complete political anthropology," but a specification of what an actor finds relevant when choosing among decisions she knows that she can take. Any verbal theory of intentional action relies on such specifications, and thus can be expressed game theoretically if one is so inclined.

For instance, if one desires a theory in which preferences change or new actors emerge only later in the game, nothing in the tools of game theory or equilibrium analysis prevents it. It is straightforward to design games in which players either are activated only after the game begins and depend on the actions of other players, or in which players present at the start of the game drop out. The former element is present in, for example, Boehmke et al. (2006), where an administrative agency can engage in policy making only if activated by an interest group; the latter is present in Gailmard and Patty (2007), where bureaucrats with weak policy interests leave the civil service. Those examples may be a more stylized form of "change" than many APD scholars might have in mind, but they accommodate change nonetheless and, thus, prove the concept that fixed and static preferences of sets of actors are not required by the technique.

In short, what game theory adds to any theory of intentional action is the clarity and consistency of formalization. It does not add a commitment to any substantive claim about theoretical content (e.g., "efficiency" of political institutions), certainty of the actors about the game being played, a static sequence of interaction, or predetermined groups of politically relevant actors.

All that said, Pierson (2004) makes a convincing case that modelers should not trumpet what their method *can* do, if none of them actually *do* it. In order to make better contributions to APD, modelers need to understand what APD scholars find inadequate or incomplete about the models presented to date, and what they are trying to accomplish in research.

1.3 Game theory and APD

How can formal theory contribute to APD? That, of course, depends on what one takes "APD" to mean. Rather than defining it textually by assigning meaning to the "A," "P," and "D," I assume that APD is what APDers do. The broadest self-definition—and the one most conducive to optimism about productive interface with formal theory—is probably that offered by the subfield's leading journal, *Studies in American Political Development*. *SAPD* defines its scope as the study of "political change and institutional development in the United States," more specifically focused on "governmental institutions over time and on their social, economic, and cultural setting." On that interpretation, APD is defined in terms of its dependent variables: change in the institutions that comprise, and policies enacted by, the American state. As long as research pertains to change and development of governmental institutions in the United States, it is APD.

Many classics of the APD genre share a focus not just on explaining change in American governmental institutions, but explaining it in a particular way. They articulate a specifically *political* logic of institutional change or state-building, i.e., a logic in which politics—in the form of ideology, electoral competition, coalition building and maintenance—is essential in itself to state building. Correspondingly, the state and its evolution are not simply epiphenomenal to more fundamental forces such as economic structure, class conflict, social cleavages, or culture, a central theoretical current in APD touchstones such as Skowronek (1982), Bensel (1990), Skocpol (1992), Sanders (1999), and Schickler (2001). Relatedly, a crucial tenet running through some seminal APD research is that the state itself and, more precisely, the actions of individuals that comprise it, is an autonomous force in institutional or policy change (or stasis); cf. Evans et al. (1985) and Carpenter (2001).

In that sense, APD is defined more restrictively: not only by its dependent variables, but also by its explanatory variables or the nature of the theory connecting them. In most of the present paper I lean toward *SAPD*'s APD-as-dependent-variable definition. That is because the issue herein is whether formal theory can contribute to formal theory in some sense, not in the most restrictive possible sense.

The potential for formal theory's contribution to APD exists for two reasons. First, APD as a *theoretical* approach and formal theory both depend on abstraction. It may seem strange to link the two research traditions on this dimension, for formal theory is (in-) famous for rarefied models detached from any specific cases, whereas APD is known for highly textured descriptions of real events. Yet the two fields share an ideal of theory-building that strips interrelationships between events down to their most essential causal forces (e.g., the constraining effect of inherited institutional structures). That reasoning process inherently also requires casting aside any factors that are not fundamental to the causal process. "Abstraction" is simply a convenient shorthand for that process of identifying the fundamental factors, building theory around them, and casting aside the rest. In neither field do causal or theoretical explanations include every facet of a decision environment faced by the decision makers under consideration. Instead, the theorist's task in both fields is to identify the facets of the decision environment that are *relevant* for explaining the events under consideration.[8] Including those relevant facets in a causal account of political action, excluding the irrelevant ones, and adjudicating between the two categories is the very purpose of a theory. No scholarship in APD strives to include every single event that occurs "in reality" in its explanation of critical decisions made by the actors being studied, because doing so would negate the clarity that can be gained from theorizing. Similarly, formal models present abstract versions of the environments, capabilities, and motivations of the decision makers being studied because it sharpens our focus on the critical factors. If interlocutors are more aware of the elements of "reality" that are excluded from a formal

[8] Indeed, the very concept of a "decision environment" is an abstraction.

model, it may be partly because models lay their assumptions bare. Theorists across the two research traditions may disagree over what constitutes relevance of some causal factor (or even proof of it). They do not disagree that theories of action should be based only on the relevant factors, and that those factors are a strict subset of all the potential factors one might enumerate. In short, the abstraction required for theory building in each field holds out a *prima facie* possibility of identifying intellectual linkages between them.

Second, both fields focus on institutions.[9] Indeed, both fields emphasize a similar duality of institutions: they are endogenous, subject to change, and in some cases the object of explanation; but from a different perspective institutions are exogenously fixed and and act to channel political action. Such duality is a major current of APD scholarship (Mettler and Valelly 2016). It also is standard fare in game theoretic modeling, wherein a specific extensive form game often is taken as a representation of a specific institution—so that institutions definitionally are the rules of the game—but institutions also may be considered as the equilibrium outcome of a game (Calvert 1995; Shepsle 1989). In the next two sections I consider formal theory's approach to that duality and contributions to APD based on it.

2 Institutions as game forms

In this section I take up the first, and predominant, view of institutions in formal political theory: institutions as extensive form games. This approach instantiates the common definition of institutions as "rules of the game" in politics and has been the approach followed in most contributions of formal theory to APD.

In the game forms approach to institutional change, a modeler considers several different game forms, to represent different institutions. The analysis then shows that the equilibria of one of the game forms are preferred to the equilibria of the other game form by an "institutional designer," a hypothetical actor empowered to structure institutional arrangements. The designer's preference counts as a sufficient explanation for the emergence of an institution because the assumption shows how some actor with the ability to create an institution had an incentive to do so.

2.1 Examples from the literature

Some of the classics of both formal political theory and the development of the US Congress adopt the game forms approach. Successive waves of this literature have focused on the committee system in the House of Representatives. In the first wave, Shepsle (1979) and Shepsle and Weingast (1981) argue from canonical social choice results that instability should be expected from collective choice in a multidimensional choice space, absent some external structure on the process. The committee system with monopoly jurisdictions, in turn, provides exactly that structure by breaking a multidimensional choice space into a sequence of unidimensional choice spaces, each of which has a Condorcet winner—and, thus, imposes a powerful centripetal restraint on collective choice. The first instance of that point (Shepsle 1979) simply recognized that a monopoly jurisdiction system would solve the "problem" without emphasizing the institution's origins or design rationale. However,

[9] In addition, though often not declared as such, much APD works in a methodological individualist framework, emphasizing the actions of individuals as constituting social outcomes. I return to that point below.

it was a short step to argue that individual members face a common interest in overcoming chaotic and unstable collective choice; thus, a developmental argument about congressional committees took hold (cf. Gamm and Shepsle 1989, wherein the “developmental” argument is about distributing turf).[10]

Shepsle and Weingast (1987) and Weingast and Marshall (1988) pushed the argument in a different direction. They argued that the committee system protects gains from trade between members in legislation (i.e., logrolls), which is especially necessary in political exchange because the trade in “goods” is not contemporaneous. One of the votes has to happen first. One cannot deny that sequence of events matters here. The committee system, it is argued, exists to facilitate such trades; it succeeds in doing so because legislators with “high demand” on a particular bundle of issues self-select onto the committees that control them, and committees generally defer to each other. Thus, “trades” between legislators are protected.

Gilligan and Krehbiel (1987) critiqued this logic and offered an alternative: committees receive deference because deference incentivized them to acquire and share information about the quality of policy alternatives on a particular issue. To be sure, committee deference allows a committee to extract some ideological rents from the legislature as a whole, but nevertheless (or rather, as a result), the committee provides a countervailing informational benefit. Gilligan and Krehbiel also embed a theory of institutional genesis in their model: in order to obtain such informational benefits, the median voter in Congress, which fully characterizes the preferences of the chamber in a one-dimensional policy space, would want to create a system of committees with deference if it did not already exist.

Another common theme of APD to which formal theory has made contributions is in the organization of bureaucracy and the executive branch. Gailmard and Patty (2007) show how the development of bureaucrats’ policy motivation, administrative expertise, long-term civil service employment, and policy discretion reinforce each other. In the “slackers and zealots” model, essentially two kinds of equilibria emerge. First is a “regime of clerkship” in which bureaucrats are not especially policy motivated, do not cultivate their expertise, are ideologically closer to their principals, enjoy little discretion, and serve short-term stints in the public sector. Second is an “expertise equilibrium” in which bureaucrats are ideologically differentiated from their principals, but nevertheless are granted discretion, acquire expertise, and build long term careers in the civil service. In essence, policy discretion is a form of compensation for bureaucrats, but only for the policy motivated ones. The model indicates how Congress can (and why it would) create bureaucratic institutions with many of the features we identify with the rational administrative state, as the cycle starts with grants of discretion by Congress, in anticipation of how bureaucrats will respond.

The institutional design process in the slackers and zealots model is run only once; the model does not explore institutional evolution. Nevertheless, imagining the model recurring over time, the two possible equilibria reveal constraints on Congress’s institutional designs. No “third type” of equilibrium—with neutral competence in the bureaucracy, or where bureaucrats invest in expertise but use it in any way Congress dictates—can exist. In essence, the model argues that the only choice in the development of administrative capacity is between a low capacity regime and a high capacity, but politicized one. It also squares a standard principal-agent logic with some of the key historical discussion in

[10] To fit the argument into a methodological individualist worldview rather than consigning it to the dreaded functionalism, it is important to identify a reason for individual members to support the institution; its collective benefits are not enough. Specifying a common interest in avoiding “chaos” does the job.

Carpenter (2001), which was at one time claimed to have eviscerated principal-agent logic in the study of bureaucratic institutions (Pierson 2004). The model thus reveals that administrative development with broad discretion and considerable ideological differentiation is consistent with a wide variety of accounts of the locus of power.

Focusing on the presidency-bureaucracy nexus, Gailmard and Patty (2012) take up the strategic foundations of the "institutional presidency." It often is observed that the president's supporting institutions provide informational advantages and help the president obtain better results in bargains with other political actors. But, if so, it is puzzling why Congress, presumably clued into that fact, continues to support those bargaining advantages for a frequent political adversary. Why not instead try to undermine the president's advantage by eliminating the institutional supports? Gailmard and Patty argue that the president's inherent discretion to act gives Congress an interest in supporting the president institutionally—even if it also means that Congress gets the short end of policy bargains sometimes. In particular, the president always has had unilateral authority to act in foreign policy and defense and, correspondingly, always has had institutional support in the bureaucracy in those areas. But in the twentieth century, the president's unilateral authority also grew with respect to domestic policy; correspondingly, Congress developed support for presidential control over the institutional presidency in those areas too. Thus, the institutional presidency has an important "supply side" component, since Congress could challenge its existence at any time, at least budgetarily. The model shows why it is in Congress's interest to instead support executive-branch institutions.

Gailmard (2017, 2019b) uses the game forms approach to study the origins of separation of powers and judicial review in the United States. The papers locate the origins in strategic problems of English imperial governance in the new world, in particular, problems for the Crown in controlling the actions of colonial governors. On separation of powers, Gailmard (2017) argues that governors with the power to be useful to a distant and weak crown also would be powerful enough to be dangerous. In particular, governors could (and sometimes did) extract taxes from colonial settlers high enough to threaten settlement and the Crown's customs revenue. The argument is that separating the colonial assembly from the governor, and endowing it with agenda setting powers in colonial taxation, could overcome this problem. In essence, the Crown liberalizes institutions to empower settlers to control governors that the Crown could not control itself.

Gailmard (2019b) notes further that empowering colonial legislatures creates its own problems with governors, namely that they might succumb to assembly pressure and approve laws against the Crown's interest. The paper argues that a forerunner of judicial review—the review of colonial legislation by the Crown in part on the grounds of consistency with English law—helped to limit opportunism by governors and induce them to withstand assembly pressure. The sequencing of institutional development, an important theme in APD (Pierson 2004), plays a role here: the problem "solved" by Crown legal review exists only because the colonial legislatures were protected from the governor's domination.

2.2 Evaluation and critique

The aforementioned papers bring attention to topics that other scholars had not considered, or had not thought about in quite the same way. For example, Gailmard (2017, 2019b) pushes temporal focus much further back than the APD's conventional concentration on the late 19th and early twentieth centuries (John 2016). Much APD has been interested in “the State” in the United States, so the English imperial era is peripheral. Still, if APD is about understanding change in US governmental institutions—or about understanding the institutions that continue to structure political conflict—then foundational institutions such as separation of powers and judicial review seem to be at least as important as defunct regulatory agencies.

On the other hand, none of the relevant literature makes significant *theoretical* innovations in the study of APD, in the sense of developing new ways of studying institutional development in time. The contributions use a standard methodology within game theory to give novel explanations for seemingly puzzling choices to cede authority or empower political adversaries. In none of them do developmental concepts such as “path dependence” or “institutional lock-in” play a role. Institutional persistence is taken as an entirely separate problem or a trivial one.[11] The separation of design and persistence is particularly stark in Gailmard (2017, 2019b) because the actors involved in design (the English Crown) are very different from those in its maintenance (American politicians).

Thus, the choice-of-game-forms perspective focuses laser-like on the specific strategic problem in which a seemingly puzzling institutional design actually is in the interest of the designer. But it offers a sharply limited and undertheorized account of institutional development. Indeed, it is not a theory of development so much as a theory of choice, and a one-time choice at that. An institution changes at a single moment (and place) in time. Why change is possible at that moment is unexplored or considered obvious, and why the product of design remains in place after that moment is not explained. APD research has focused some attention on that issue (e.g., Schickler 2001), but formal modeling has not contributed to the discussion. In that sense, the game forms approach to institutional development shares much in common with “critical juncture” explanations commonly invoked in historical sociology (Pierson 2004, ch. 5). The “institutional design” window is open at a particular moment, but the reasons are given exogenously; once a choice is made, the window closes and stays closed for hazily specified reasons.

Moreover, the reason that the “institutional design” problem came down to a single choice by a single powerful actor also often is not explained in formal models of choice of game forms.[12] From a developmental perspective, the choice of pivotal actor is in some respects the most important problem, but the theories often are silent on it. Probably most APDers would agree that *if* institutions are “designed” by a single powerful actor, they will be designed to suit that actor’s strategic interests. The key question is not necessarily the subtle details of that strategic interest, on which modelers usually focus, but the configuration of events that led to a singular moment of institutional design in the first place.

Historical institutionalists considering the institutions-as-game-forms methodology have named it “actor-centered functionalism” (Pierson and Skocpol 2002; Pierson 2004), or ACF. ACF holds that “institutions take the form they do because because powerful actors

[11] In some of the models summarized above, it is not so much that a design window “closes,” as that the designer is presumed to face the same design incentives every time it is open. In that sense, the problem of institutional “persistence” is solved trivially, with no sense of institutional change after the initial design is in place.

[12] In some of cited models, a constitutional reason is found to focus on a specific designer. Or a background analytical result identifies one in particular out of a larger constitutionally-specified set. For example, in studies of congressional rules, Article I section 3 of the US Constitution states that each chamber shall choose its own procedures. From there, the median voter theorem, combined with an assumption of a unidimensional policy space, implies that the median of the chamber is decisive.

engaged in rational, strategic behavior are seeking to produce the outcomes observed" (Pierson 2004, p. 14).[13]

Pierson critiques ACF on several grounds, including:

1. Institutions may have multiple effects, with different ones important to different members of a designing coalition.
2. Pressures of political survival often induce a short run orientation (i.e., a small discount factor), so long-term consequences may not play much role in institutional design.
3. Institutional effects may be unanticipated, in which case they necessarily cannot factor into institutional design.
4. Institution builders may respond to non-instrumental motives, e.g. a "logic of appropriateness" rather than a "logic of consequence."
5. Designer continuity

For those reasons, it is hazardous to observe long-run effects and infer that they were intended in a singular time-bounded act of institutional design. The points are important, but much of the literature cited above is not based on such an inference. For instance, in Gailmard (2017, 2019b), the long-term effects of separation of powers and judicial review do not factor into the theory of institutional design by the Crown.[14] More generally, it is straightforward to build models that are not subject to the mentioned critiques. For instance, defining the set of "outcomes" to include individual actions allows for payoffs to inhere in choices themselves, which in turn captures one effect of social norms or "appropriateness" in a standard consequentialist framework. Yet all of this still falls clearly within an ACF approach.

With outcomes and the institution designer's time horizon properly specified, the remaining critiques of ACF can seem baffling. After all, no one would want to suppose that some actor with the authority to design an institution would ignore its effects on outcomes that actor cares about, or knowingly would choose an institution that produces worse effects.

But the fundamental theoretical blindspot in ACF is not its assumption of consequentialism, rationality, or strategic foresight over some (possibly short) time horizon. It is that politics in those models unfolds invariably on a stage designed at a specific point in time by a few powerful individuals. It is not clear how to theorize about the accretion of actions by a large number of small individuals into an "institution" in the ACF-based, game forms approach. Surely that approach is not always mistaken; focus on institutional change caused by pivotal "entrepreneurs" or "skilled social actors" is common in APD and historical institutionalism as well (Pierson 2004, pp. 141–143; Sheingate 2003). But if we should view politics as a realm in which the many act within the structures created by a powerful few, that view should be a considered choice against alternative explanations, not accepted implicitly because the theoretical tool forces it.

[13] We might quibble that the characterization is not functionalism, a label that no methodological individualist would self-apply. But it largely is beside the point: if we gave that style of analysis some other name, all of the same critiques would apply.

[14] Indeed, the separation of powers paper implicitly assumes that the downsides of empowering legislatures, which necessitated Privy Council legislative review, were not anticipated by the Crown, even though they materialized within a few decades.

Put differently, institutions “should often be seen as the *by-products* of social processes rather than the realization...of actors’ goals” (Pierson and Skocpol 2002). To formal theorists, a rough and ready isomorphism to “institutions are a by-product of social processes” is “institutions are the outcome of a game.” That is, á la Shepsle (1989) and Calvert (1995), formal theorists are not bound to think of institutions exclusively as game forms; they also think of institutions as outcomes in themselves. That approach is considered in the next section.

3 Institutions as equilibria

In the second major strand of formal theorizing relevant to APD, institutions are understood as equilibria, not as game forms, adopting the perspective of “institutions” as stable, repeated patterns of behavior. In that approach, a game (often an infinitely repeated game) is specified in which a variety of patterns of behavior can be contemplated. The game embeds incentives in which a particular pattern emerges as an equilibrium institution. No one actor dictates the equilibrium reached; equilibrium is an emergent property of the system of interaction. Thus, the approach emphasizes institutions as a byproduct of unplanned, uncoordinated action by individuals. Such theories also inherently deliver a theory of institutional persistence, because equilibria are self-enforcing.

3.1 Examples from the literature

Perhaps the earliest example in the literature with implications for APD is McKelvey and Riezman (1992) on the “seniority system” in Congress. McKelvey and Riezman embed an infinitely repeated “divide the dollar” game in a model of electoral accountability. Voters reelect their member (or not) based on expectations of future benefits claimed for their district. The stationary equilibrium exhibits a “seniority system” in which legislators claim larger benefits after their first period of service. Voters will reelect a junior member who produced few benefits because they expect that member to obtain larger benefits in the future; replacing a rising senior member with a new junior member merely prolongs the district’s wait until more largesse is received. On a related theme but using a more empirically tractable model, Kanthak (2011) examines how the seniority system is developed and maintained by legislators’ retirements when legislative rules do not favor them.

Some of the now-classic pieces on American institutions as equilibria emphasize the self-enforcing character of institutions, that is, focus on the question of institutional persistence. For example, Defigueiredo and Weingast (2005) analyze the self-enforcing properties of US federalism. States decide how much to centralize power, which is beneficial because of scale effects in public policy, but potentially costly because a strong central government can take advantage of individual states, extracting wealth from them. Defigueiredo and Weingast focus on trigger equilibria of a repeated game in which states contribute to the center every period, the center resists temptation to extract rents and the scale benefits of public policy are realized. Federalism with those contours is self-enforcing in the trigger equilibrium: if the center ever expropriates from a state, the states punish it forever with small contributions.[15]

[15] See also Weingast (1997) on self enforcing rule of law in the shadow of public protest, Dragu and Polborn (2013) on self enforcing rule of law when administrators concerned about future punishment are the

Larson (2017) uses a similar methodology (i.e., trigger strategies in a repeated stationary game) to study governance with a weak state and ethnic heterogeneity, with applications to social norms in the nineteenth century American West. Heterogeneity is represented by an explicit network structure. Larson shows that little or no overlap of an ethnic minority group with a predominant social network supports strictly less cooperation in equilibrium, particularly between members of the predominant group and the minority. On that basis she explains the erosion of social norms and discrimination after rapid immigration of Chinese enclaves in Western mining areas.

A different approach is taken by Gailmard (2019a) to analyze the evolution of legislative power in Britain's American colonies. Here, the legislature's power emerges as a byproduct, rather than an intention, of the Crown, and it emerges gradually over multiple periods. The Crown grants institutional power to the governor, its agent, but the legislature would like to claim that power for itself. The governor has private information about his "resolve" in resisting legislative challenges. The key tension is that when that resolve is weak, the governor wishes to capitulate to demands for power from the colonial legislature, but not to reveal that capitulation to the Crown. The governor exploits random noise in the policy making environment to make gradual concessions in equilibrium, which are indistinguishable by the Crown from the actions of a strong governor. But once a concession is observed by the legislature, more challenges (and concessions) are forthcoming in the future, so the "institution" of legislative power changes gradually over time in equilibrium.

3.2 Evaluation and critique

Most of the models above emphasize the self-enforcing character of the equilibrium. Many equilibria exist in those games, and little or no attention is paid to the selection of a particular one.[16] Nor does the institution-as-equilibrium change over time: all such cases involve repetition of the equilibrium outcomes. If the per-period equilibrium outcome changes, it is an instantaneous reaction to a short-term exogenous parameter shift (e.g., change in network structure).

As with the institutions-as-game-forms approach, no extensive *theoretical* innovation in the institutions-as-equilibria approach has emerged in the past 25 years. The contributions to the literature emphasize substantive advances illuminated using a standard (among modelers) methodology. They do not theorize about the problems of institutional development or persistence, or represent the time horizon of institutional change, in new ways.

That state of affairs simply recognizes the point of Greif and Laitin (2004), that game theoretic models thus far have not generally been built to explore a nontrivial dynamic unfolding of institutional change. The exception to that conclusion among the papers cited is Gailmard (2019a), in which institutional change unfolds gradually over time and a nontrivial transition dynamic occurs in equilibrium. However, expressing dynamic features of institutional evolution comes at the cost of tractability. The model contains only three periods; the others cited are repeated infinitely. In addition, the paper does not propose a

Footnote 15 (continued)

ones who carry out malfeasance, and Fearon (2011) on self enforcing democracy in the sense of regular contested elections.

[16] When stationarity is applied, a loose justification usually is provided on grounds of behavioral simplicity. With trigger strategies (which do not sustain stationary equilibria), a loose Pareto efficiency rationale for equilibrium selection often is offered.

general theoretical approach to understanding gradual institutional change; it develops a model very closely tailored to a specific situation in which an important change occurred. On the other hand, it is not clear that progress on understanding institutional dynamics in specific contexts is better approached by attempting to distill an abstract essence of institutional change that is independent of context.

4 New directions and critiques from APD

In summary, formal theorists have taken two approaches to studying institutional change, development, or persistence in American politics. In the first, institutional design is modeled as a choice of game forms by a specific, powerful actor in the political process. In the second, institutions are conceived as equilibria of a more fundamental game, sustained by mutually reinforcing behavior of multiple actors. Thus, institutions are not designed; they are the byproducts of social interaction.

The first approach has been used to study institutional change and emergence, albeit in a highly stylized form of a short term (within a single game period) design problem. But that approach usually takes the long-term stability or persistence of institutions as given, or assumes away the problem of persistence by imagining repetition of the same design problem with the same designer.

The second approach has been used primarily to study institutional persistence, i.e., long-term stability. But in the equilibria of infinitely repeated games, the per-period play typically is constant. That approach has been adopted only rarely to study institutional change. Of course, it is possible to build a single model with both a singular "institutional design" phase and a repeated game "institutional maintenance" phase. That is exactly the approach of Defigueiredo and Weingast (2005) on federalism and of Buchanan and Tullock (1962) on constitutional political economy.[17] Nevertheless, design and persistence remain as two conceptually separate processes.

The methodology for those approaches is relatively well established. Theoretical innovations in studying institutional development, change, and persistence certainly are occurring (e.g., Acemoglu and Robinson 2000; Penn 2009; Callander and Hummel 2014; Acemoglu et al. 2015; Bednar and Page 2018), though not thus far in the literature on American institutional development. It would be useful to think about APD from such perspectives.

What modelers must guard against in doing so is taking an existing technique off the shelf and searching through history for an apt case, what Pierson and Skocpol (2002) call "illustrative history," or "the mining of the historical record for outcomes which can be 'explained' by particular rational choice models." Such an approach gives short shrift to understanding substantive institutional developments almost by construction; as such, it is not a useful way for formal theory to contribute to APD.

New modeling approaches aside, it is fair to say that not all scholars of institutional change and development are equally sanguine about the possibility of useful pivots in formal theorizing on those topics. Scholars of APD and the related field of historical institutionalism have critiqued the prospects for *any* formal, game theoretic model of institutional development on several grounds. I will consider some of them in turn.

[17] That approach is also reminiscent of "state of nature" philosophers and the biblical Garden of Eden myth.

4.1 Game theory and temporal sequences

Pierson (2004) contends that several problems plague the application of game theory to the study of temporal sequences. Inasmuch as those sequences are important in, if not constitutive of, institutional development, that claim suggests major limits on the ability to contribute to the study of institutional development with game theoretic analysis. In particular, Pierson (2004, pp. 82–92) argues that both the causes and the consequences of institutional change may unfold slowly over long periods of time. Many modes of analysis implicitly slant toward short run orientations on both dimensions, but in the process overlook important aspects of change, e.g., threshold effects and path dependence. It is fair to say that such neglect has been characteristic of formal modeling in APD, suggesting that modelers interested in APD should take a much broader view of institutional development in time. Doing so will tap into both interesting frontiers in formal theory and find an interested audience among APD scholars.

Beyond that, Pierson (2004, pp. 61–62) points to four specific limitations of game theoretic modeling of social processes, which I consider in turn with some comments.

1. "Game theory itself can say nothing about payoffs and preferences," i.e., it must take them as given. That is true in a technical sense, but not in a substantive one. Flexible models of preference change can be obtained by adding new information and belief updating (standard fare in the literature since Gilligan and Krehbiel 1987), or by incorporating endogenous evolution of which players play which roles in the game. For just one example, in Gailmard and Patty (2007) the average preferences of "bureaucrats," in the sense of the people who actually work for the government's administrative agencies, evolve over time, even though the preferences of potential bureaucrats do not.

   Moreover, for studying institutional development, some of the most important preference changes are those induced over actions, not fundamental preferences over outcomes. For example, with path dependence in modes of health care delivery, one is not postulating change in preferences over outcomes such as "health" or "cost of achieving a given level of health." One is postulating that the *induced* preference for a mode of health care delivery at date t depends on the mode adopted at $s < t$, perhaps because it is costly to switch modes, implying only that, for a given evaluation of switching costs (i.e., a fixed preference), one is less inclined to do it. Thus, change in relevant policy or action choices can easily be accommodated without change in preferences over outcomes.
2. "Game theory needs to focus on relatively cohesive, well-integrated 'composite actors'.... It has great trouble integrating 'quasi-groups'," collections of individuals that cannot be treated as unitary strategic actors but whose "utility functions are interdependent in such a way that certain acts by some will increase or decrease the likelihood that others will act in the same way." That critique contains more than some truth. It is common for formal modelers to speak of "an agency" or "the legislature," even "the poor" or "the elites." In some cases, that is a convenient synecdoche on top of an analytical result justifying it (e.g., the median voter theorem for collective preferences, the Meltzer-Richard model of majoritarian redistributive preferences). In other cases, such actors indeed are treated as unitary decision makers ("a highly questionable move" as Pierson and Skocpol (2002) put it).

   On the other hand, quasi-groups are relatively common in APD scholarship as well; one is not surprised to read of actions by "the Republican Party" or "Northern financiers" or "the Grangers" or "Congress." It is not clear what if any method is available in this literature to avoid the problem of quasi-groups. What, precisely, is the way in which the utility functions of quasi group members become interdependent, such that common actions propagate within the group? If interdependence and common actions simply are assumed verbally or sidestepped altogether, it is no better than treating the group as a unitary actor formally. But if that process can be described in words, it can be described in symbols as well. Such an exercise for its own sake is useless, but if nothing else formalization will clarify when we have no theory of how quasi-group members come to exhibit common preferences and take common actions. It cannot be hidden in a model.

3. "Games need to be kept very simple: few actors, few options." That is perhaps a matter of taste. Larson (2017) explicitly discusses a special case of her model with no fewer than 206 players, which seems like a lot. It seems likely that the typical configuration with a small number of players and actions, along with "short" sequences, stems from a preference for parsimony. That preference seems reasonably common among APD scholars as well. For example, Bensel (1990) gives a sophisticated and subtle account of the failure of Reconstruction by casting only Northern Republicans, Northern "financial capitalists" (a composite actor), and readmitted Southern Democrats in the starring roles.
4. "Sequences cannot be interrupted. Sequence, in these models, refers to an ordered alternation of 'moves' by 'composite actors'..." That is true, but may partly be a definitional quibble. If one wants to consider possible "interruptions" of a sequence, one defines a new sequence with the possible interruption modeled. The chief limitation here may be that all such interruptions must be foreseeable, in which case the core issue is the assumption noted above that unforeseen contingencies are at present beyond the reach of game theory.

Doubtless, formal theorists and their critics could entertain themselves endlessly with volumes full of spirited critiques, rejoinders, and clarifications. Perhaps a more productive approach is for scholars to engage directly with the explanations offered for institutional development in individual pieces of scholarship, formal or otherwise. Then, they should highlight cases of excessive simplicity (or complexity, or hidden assumptions) for discussion among interested colleagues. In that way, explanations and our collective state of knowledge can improve. Ideally, such a dialogue could take place across methodological persuasions, but doing so places a high premium on good communication.

4.2 Individualism, methodological and otherwise

One such specific critique of formal modeling in APD focuses on the level of aggregation in institutional analysis. APD scholars and fellow travelers argue that formal modelers have an excessively "individualist" focus. For example, Orren and Skowronek (2004, p. 18) hold that institutions, while "by no means neglected" in rational choice theory, are "subordinated by theories and methods that are essentially individualist, and their role correspondingly attenuated." Similarly, Pierson and Skocpol (2002) contend that

> Rational choice practitioners typically focus on political contexts with coherent strategic actors—preferably individuals, such politicians or candidates—operating in particular, well-bounded contexts—such as legislatures where choices are clearly identifiable and payoffs relatively transparent. Efforts to deal with broader

> social aggregates, whether interrelated organizations or looser social groupings, are often avoided.

By contrast, APD has a "bias in favor of the polity as a whole," which is

> a significant departure in a field where the standard curriculum is divided into operation of the parts...the branches of government, the parties, the interest groups, the electorate, and so on separately. In this respect, APD is different also from the 'new institutionalism' of rational choice, where modeling techniques likewise focus analysis on the behavior of actors in particularized settings of rules or game forms (Orren and Skowronek 2004, p. 185).

Similarly, Pierson and Skocpol (2002) note that, "arguably, instead of the intellectual problems faced by rational choice getting bigger, the universe of politics deemed as suitable for scrutiny gets redefined in ever more diminutive terms. Thus the study of American politics becomes the study of Congress (or, at its most expansive, the study of Congress and administrative agencies)."

It is tempting to read critiques of "essentially individualist" approaches in formal modeling as critiques of methodological individualism, which is baked into game theory (the modeling of "composite actors" as unitary decision makers notwithstanding). I believe that that would be a misinterpretation. To be sure, APD and the closely related historical institutionalism aspire to address macro-social causal explanations and contexts (Pierson and Skocpol 2002). And, given the theoretical diversity in those fields, plenty of scholarship of course adopts "organizational," evolutionary, even functionalist, or otherwise non-methodologically individualist approaches. Nevertheless, a focus on individual decisions—an "action frame of reference" that underpins methodological individualism—is, while not a defining feature of APD, very common. Classics such as *Building a New American State, Protecting Soldiers and Mothers, Yankee Leviathan, Roots of Reform, The Forging of Bureaucratic Autonomy*, and *Disjointed Pluralism* all focus on tectonic movements of institutions over decades, but they do so by drilling down into the decisions of individuals (cf. Carpenter 2003). One becomes closely acquainted with the proper names of specific actors who had goals, faced constraints, and took decisions. The theoretical content of those works is not in the inexorable march of macro-scale, temporally-unbounded socioeconomic aggregates—the Proletariate, the Church, the Capitalists, the Public. Even "the State," having been brought back in, often is disaggregated in such works into individuals who have interests, beliefs, and constraints. Things do not happen in those analyses simply because The Institution needed them to happen.

Instead, at least part of the critique of formal theoretic "individualism" in APD and historical institutionalism seems to be about what we might call "institutional atomism": carving the US government or even the whole political system into component parts, and then explaining each component individually, perhaps mostly with reference to internal workings of that component. For example, institutional atomism would hold that Congress can be understood entirely by focusing on the internal proceedings of Congress—or even one single chamber.

Naturally, the optimal set of explanatory factors for some dependent variable (such as the House's committee structure) is an empirical question. Still, it seems wise for formal theorists to heed the critique. First, since political actors aim to influence outcomes such as policy enactments or office holding that extend beyond any one institution, it simply seems unlikely that most important institutions in American politics can be explained adequately in an institutionally atomistic framework. Second, expanding focus beyond isolated pockets will keep models focused on interesting and probably more important questions.

4.3 Suspense, contingency, and conjuncture

In his critique of *Analytic Narratives*, Carpenter (2000) argues that interpreting history through finite, extensive form games of complete and perfect information tends to suppress if not eliminate some key elements of historical explanation generally and narrative specifically:

- Suspense: key developments are not foreordained
- Contingency: multiple possible histories are possible
- Conjuncture: simultaneous occurrence of multiple processes that together trigger a critical event

Formal theorists interested in APD should heed the call to foreground those elements, not elide them when constructing models. It is important to note that Carpenter's critique is trained carefully on the use of a specific type of game. In other classes, it is straightforward to include formal elements that produce the desired features. In games of incomplete or imperfect information, the role of chance (either exogenous or endogenous) is prominent, and it is difficult to maintain that a given path of play is foreordained—even within a given equilibrium.

Similarly, a sense of contingency is inherent in games with multiple equilibria. Modelers often treat such "indeterminacy" as a bug, but it could equally well be seen as a feature. Rather than refining away the undesired equilibria—or changing the game form to eliminate them—modelers can embrace contingency by exploring the process of coordinating on a specific one. That approach dovetails naturally with a sense of "lock-in," since equilibria, once coordination is achieved, are self-enforcing.

Conjuncture can be represented formally in several ways; a natural one is with stochastic games. In that class of games, "states" occur at random that can change the payoffs or the game's extensive form. Such states can represent exactly the critical conditions such as war, drought, or social unrest that can alter dramatically incentives of other players to make irreversible decisions. That is the approach taken, for example, by Acemoglu and Robinson (2000) in the analysis of elite incentives to democratize.

The foregoing merely is a suggestive, not exhaustive, account of how models might incorporate important elements of historical processes. The point is simply that incorporating them is possible, and the principal requirement for doing so is awareness that it would be desirable. Formal models of APD have not thus far taken those approaches, but it would be useful for connecting with themes that are important in APD scholarship.

5 Conclusion

Formal theorists who wish to study American political development can very well go about their modeling exercises with no concern for how APDers evaluate their work. Nobody "owns" history and any scholar can analyze it as she pleases. However, such bifurcation is suboptimal both for those interested formal theorists and for the field of APD. First, it limits the critiques that the modelers will hear and heed. In such a pattern, modelers will talk to modelers who can critique models, but neither the substance nor the sensitivity to theoretical themes in which modelers do not at present excel. Second, it limits the density of the network of APD scholars who are in dialogue with one another, and it limits the range of theoretical perspectives to which any camp of interested scholars is exposed. Since that density is not at present so large that scholars need to fragment in order to have useful conversations, such fragmentation undermines the field as a whole.

If productive dialogue between formal theory and APD is to occur, then each camp will need to better understand the other. For formal theorists to understand requires fuller appreciation of important critiques from APDers and historical institutionalists. Those include critiques of using history as an illustrative example, and studying institutional *choice* versus institutional *development*. A minimal antidote for the former problem is to read some history first, then develop a model. Going in the other direction almost invariably leads to reading history in search of examples that conform to a model, which is nothing but confirmation bias. No scholar serious about understanding the historical case at hand can be faulted for giving short shrift to such theory.[18]

Formal theorists have made some notable contributions to APD—some theoretical (especially on self-enforcing institutions), many substantive. Yet we are barely scratching the surface of the potential innovations in formal theory suggested by institutional development, or of potential understandings of institutions afforded by formal models. In order to realize that potential, formal theorists much accept the obligation to communicate better about the contents of their models, and to rise to critiques of the types of institutional change and development embedded in models thus far. Only by understanding those critiques better can formal theorists communicate in such a way that their audience does not try to define them out of the conversation.

Acknowledgements Thanks to Jeff Jenkins, Kristin Kanthak, John Patty, and participants at the USC Conference on Causal Inference and APD for helpful comments.

References

Acemoglu, D., Egorov, G., & Sonin, K. (2015). Political economy in a changing world. *Journal of Political Economy*, *123*(5), 1038–1086.

Acemoglu, D., & Robinson, J. A. (2000). Why did the west extend the franchise? Democracy, inequality, and growth in historical perspective. *The Quarterly Journal of Economics*, *115*(4), 1167–1199.

Ballingrud, G., & Dougherty, K. L. (2018). Coalitional instability and the three-fifths compromise. *American Journal of Political Science*, *62*(4), 861–872.

Bates, R. H., Defigueiredo, R., & Weingast, B. (1998). The politics of interpretation: Rationality, culture, and transition. *Politics & Society*, *26*(4), 603–642.

Bates, R. H., Grief, A., Levi, M., Rosenthal, J.-L., & Weingast, B. (1998). *Analytic narratives*. Princeton, NJ: Princeton University Press.

Bednar, J., & Page, S. E. (2018). When order affects performance: Culture, behavioral spillovers, and institutional path dependence. *American Political Science Review*, *112*(1), 82–98.

[18] I am not suggesting a simplistic separation of first reading history with no theoretical priors, and then developing a model without any further investigation of history. I am suggesting that a simplistic separation of modeling first, and engaging the substantive history only after that is complete, is unlikely to produce a compelling account. I also am aware that my recommendation contradicts the "middle school science" model in which one cannot theorize after analyzing any data.

Bensel, R. F. (1990). *Yankee Leviathan: The origins of central state authority in America, 1859–1877*. New York: Cambridge University Press.

Boehmke, F. J., Gailmard, S., Patty, J. W., et al. (2006). Whose ear to bend? Information sources and venue choice in policy-making. *Quarterly Journal of Political Science*, *1*(2), 139–169.

Buchanan, J., & Tullock, G. (1962). *The calculus of consent*. Ann Arbor, MI: University of Michigan Press.

Callander, S., & Hummel, P. (2014). Preemptive policy experimentation. *Econometrica*, *82*(4), 1509–1528.

Calvert, R. L. (1995). Rational actors, equilibrium, and social institutions. In J. Knight & I. Sened (Eds.), *Explaining social institutions* (pp. 57–95). Ann Arbor, MI: University of Michigan Press.

Carpenter, D. (2000). Commentary: What is the marginal value of analytic narratives? *Social Science History*, *24*(4), 653–667.

Carpenter, D. P. (2001). *The forging of bureaucratic autonomy: Reputations, networks, and policy innovation in executive agencies, 1862–1928*. Princeton, NJ: Princeton University Press.

Carpenter, D. P. (2003). The multiple and material legacies of stephen skowronek. *Social Science History*, *27*(3), 465–474.

Defigueiredo, R., Rakove, J., & Weingast, B. R. (2006). Rationality, inaccurate mental models, and self-confirming equilibrium: A new understanding of the american revolution. *Journal of Theoretical Politics*, *18*(4), 384–415.

Defigueiredo, R., & Weingast, B. R. (2005). Self-enforcing federalism. *Journal of Law, Economics, and Organization*, *21*(1), 103–135.

Downs, A. (1957). *An economic theory of democracy*. New York: Harper.

Dragu, T., & Polborn, M. (2013). The administrative foundation of the rule of law. *The Journal of Politics*, *75*(4), 1038–1050.

Evans, P. B., Rueschemeyer, D., & Skocpol, T. (1985). *Bringing the state back in*. New York: Cambridge University Press.

Fearon, J. D. (2011). Self-enforcing democracy. *The Quarterly Journal of Economics*, *126*(4), 1661–1708.

Gailmard, S. (2017). Building a new imperial state: The strategic foundations of separation of powers in America. *American Political Science Review*, *111*(4), 668–685.

Gailmard, S. (2019a). British imperial governance and the development of legislative power in America. *UC Berkeley Typescript*.

Gailmard, S. (2019b). Imperial politics, english law, and the strategic foundations of constitutional review in America. *American Political Science Review*, *113*(3), 778–795.

Gailmard, S., & Patty, J. W. (2007). Slackers and zealots: Civil service, policy discretion, and bureaucratic expertise. *American Journal of Political Science*, *51*(4), 873–889.

Gailmard, S., & Patty, J. W. (2012). *Learning while governing: Expertise and accountability in the executive branch*. Chicago: University of Chicago Press.

Gamm, G., & Shepsle, K. (1989). Emergence of legislative institutions: Standing committees in the house and senate, 1810–1825. *Legislative Studies Quarterly*, *14*, 39–66.

Gilligan, T. W., & Krehbiel, K. (1987). Collective decisionmaking and standing committees: An informational rationale for restrictive amendment procedures. *Journal of Law, Economics, & Organization*, *3*(2), 287–335.

Greif, A., & Laitin, D. D. (2004). A theory of endogenous institutional change. *American Political Science Review*, *98*(4), 633–652.

Jenkins, J. A. (2016). APD and rational choice. In R. Valelly, S. Mettler, & R. C. Lieberman (Eds.), *The Oxford handbook of American political development* (pp. 148–165). New York: Oxford University Press.

John, R. R. (2016). American political development and political history. In R. Valelly, S. Mettler, & R. C. Lieberman (Eds.), *The Oxford handbook of american political development* (pp. 185–206). New York: Oxford University Press.

Kanthak, K. (2011). The hidden effects of rules not broken: Career paths, institutional rules and anticipatory exit in legislatures. *British Journal of Political Science*, *41*(4), 841–857.

Larson, J. M. (2017). Why the west became wild: Informal governance with incomplete networks. *World Politics*, *69*(4), 713–749.

Maskin, E., & Tirole, J. (1999). Unforeseen contingencies and incomplete contracts. *The Review of Economic Studies*, *66*(1), 83–114.

McKelvey, R. D., & Riezman, R. (1992). Seniority in legislatures. *American Political Science Review*, *86*(4), 951–965.

Mettler, S., & Valelly, R. (2016). Introduction: The distinctivenss and necessity of american political development. In R. Valelly, S. Mettler, & R. C. Lieberman (Eds.), *The Oxford handbook of American political development* (pp. 1–26). New York: Oxford University Press.

Miller, G., & Schofield, N. (2003). Activists and partisan realignment in the United States. *American Political Science Review*, *97*(2), 245–260.

Orren, K., & Skowronek, S. (2004). *The search for American political development*. New York: Cambridge University Press.

Penn, E. M. (2009). A model of farsighted voting. *American Journal of Political Science*, *53*(1), 36–54.

Pierson, P. (2004). *Politics in time: History, institutions, and social analysis*. Princeton: Princeton University Press.

Pierson, P., & Skocpol, T. (2002). Historical institutionalism in contemporary political science. In I. Katznelson & H. Milner (Eds.), *Political science: The state of the discipline* (pp. 693–721). New York: W. W. Norton.

Sanders, E. (1999). *Roots of reform: Farmers, workers, and the American state, 1877–1917*. Chicago: University of Chicago Press.

Schickler, E. (2001). *Disjointed pluralism: Institutional innovation and the development of the US Congress*. Princeton, NJ: Princeton University Press.

Schofield, N. (2006). *Architects of political change: Constitutional quandaries and social choice theory (Political Economy of Institutions and Decisions)*. New York: Cambridge University Press.

Sheingate, A. (2003). Political entrepreneurship, institutional change, and american political development. *Studies in American Political Development*, *17*(2), 185–203.

Shepsle, K. A. (1979). Institutional arrangements and equilibrium in multidimensional voting models. *American Journal of Political Science*, *14*, 27–59.

Shepsle, K. A. (1989). Studying institutions: Some lessons from the rational choice approach. *Journal of Theoretical Politics*, *1*(2), 131–147.

Shepsle, K. A., & Weingast, B. R. (1981). Structure-induced equilibrium and legislative choice. *Public Choice*, *37*(3), 503–519.

Shepsle, K. A., & Weingast, B. R. (1987). The institutional foundations of committee power. *American Political Science Review*, *81*(1), 85–104.

Skocpol, T. (1992). *Protecting soldiers and mothers*. Cambridge, MA: Harvard University Press.

Skowronek, S. (1982). *Building a new American state: The expansion of national administrative capacities, 1877–1920*. New York: Cambridge University Press.

Weingast, B. R. (1997). The political foundations of democracy and the rule of the law. *American Political Science Review*, *91*(2), 245–263.

Weingast, B. R., & Marshall, W. J. (1988). The industrial organization of congress; or, why legislatures, like firms, are not organized as markets. *Journal of Political Economy*, *96*(1), 132–163.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.