Formal Models of Nondemocratic Politics

Scott Gehlbach, Konstantin Sonin, and Milan W. Svolik

1Department of Political Science, University of Wisconsin–Madison, Madison, Wisconsin 53706; email: gehlbach@polisci.wisc.edu
2Harris School of Public Policy Studies, University of Chicago, Chicago, Illinois 60637; email: ksonin@uchicago.edu
3Higher School of Economics, Moscow 101000, Russia
4Department of Political Science, Yale University, New Haven, Connecticut 06511; email: milan.svolik@yale.edu

Keywords
dictatorship, autocracy, regime change, institutions, electoral manipulation, media control, formal political theory, political economy

Abstract
The last decade has witnessed growing interest among political scientists and economists in nondemocratic politics. This trend has been reflected in increasingly rigorous game-theoretic modeling of its various aspects: regime persistence and breakdown, ruling-coalition formation and leadership change, protests and repression, formal institutions and elections, and censorship and media control. We review this research agenda, focusing on the foundational assumptions and political intuition behind key models. Our survey reveals a field populated by disparate models of particular mechanisms that nonetheless share two major analytical themes: asymmetric information and commitment problems. We propose that future models move toward a genuinely comparative study of authoritarian institutions.
INTRODUCTION

Dictatorship is the most prevalent form of government in history. Yet only in the last decade have political scientists and economists begun to examine nondemocratic politics with the same empirical scope and theoretical rigor that they have traditionally devoted to the study of democratic politics. This trend has been reflected in the development of increasingly sophisticated game-theoretic models of nondemocratic politics, including models of regime and leadership dynamics (regime persistence and breakdown, the nature and frequency of leadership changes, ruling-coalition formation), authoritarian institutions (parties, legislatures, elections), and policies (repression, censorship, cooptation). In this article, we review important contributions to this research agenda and critically examine the foundational assumptions and political mechanisms at work in key models.

The study of authoritarian politics presents both challenges and opportunities for formal political theory. The challenges stem from a fundamental distinction between dictatorship and democracy. In democracies, power is by definition contested via free and fair elections, and political conflicts are resolved institutionally; polities for which this is not the case are considered dictatorships. In turn, democratic politics is particularly amenable to formal modeling, as key features of a democracy’s constitutional framework, such as electoral or legislative rules, often delineate precisely those elements that are required for a well-defined game-theoretic model: the players, their available actions and information, and the procedures by which actions translate into outcomes.

By contrast, which institutions, rules, or even leaders govern in dictatorships is often unclear or contested. This is because a great deal of authoritarian politics either is noninstitutional and entails the threat or overt use of violence—as in the case of coups, protests, and repression—or is aimed at circumventing or subverting formal institutional rules—as in the case of electoral fraud and manipulation. In fact, a regime’s ability to exercise powers beyond any constitutional constraints is often seen as the essence of dictatorship.

When studying nondemocratic regimes, therefore, we cannot take formal institutions at face value. This does not automatically imply that formal institutions in dictatorships are inconsequential. But it does raise methodological challenges that we less frequently encounter in the study of democratic politics. Most urgently, we must explain not only the equilibrium consequences of institutions, but also why—in light of those consequences—political actors have an incentive to comply with institutions in the first place. More generally, realistic models of authoritarian politics must recognize that institutional rules may be circumvented, that political conflicts may be resolved violently, and that information is often limited or asymmetric.

Formal models of authoritarian politics that account for these analytical challenges often make contributions that go beyond an improved understanding of nondemocratic regimes alone. When we analyze institutions as equilibria, for instance, the often-emphasized point that “institutions are endogenous” arises directly out of our models. Formal analysis of authoritarian institutions thus helps us understand the challenges that the endogeneity of institutions presents for the empirical evaluation of the role and consequences of institutions across political regimes.

1We use the terms dictatorship, autocracy, authoritarian regime, and nondemocratic regime interchangeably.

2Svolik (2012) finds that about two-thirds of all leadership changes in dictatorships between 1946 and 2008 were noninstitutional (i.e., they violated official rules or established conventions), almost one-half involved the military, and about one-third entailed overt violence.

3See Przeworski (2009) and Pepinsky (2014) for a discussion and Acemoglu et al. (2001) for an influential attempt to overcome concerns about the endogeneity of institutions.
Another contribution of formal models of authoritarian politics that extends beyond their immediate focus is heuristic. A valuable by-product of the abstraction and simplification involved in formal modeling is an analytical transparency that highlights the generality of key political mechanisms in models of authoritarian politics. This heuristic value is most apparent in areas of political science and economics that may have a substantively different focus yet share some of the same conceptual challenges that arise in the modeling of nondemocratic politics: the questionable relevance of formal institutions, the absence of a higher authority to enforce contracts, the prominent role of violence in the resolution of political conflicts, and pervasive asymmetries of information.


Finally, formal models of authoritarian politics highlight that differences in behavior between dictatorships and democracies need not be driven by differences in the preferences of dictators and democrats. Perhaps the key assumption in the models we review is that autocratic rulers are motivated by the same desires—political survival, ideology, rents—as their democratic counterparts, and that they are similarly rational in maximizing their preferences. The crucial difference between dictators and democrats is that the former are much less constrained in how they can pursue their goals, implying a broader range of means to these universal ends. As we discuss at length below, the field has responded to this empirical richness with a diversity of models that capture numerous stylized nondemocratic environments.

Conceptually, our review starts with some of the most distinctively nondemocratic topics: ruling-coalition formation, revolution, and repression. We next consider the role of formal institutions in nondemocratic regimes. Finally, we examine authoritarian elections and media control and the attendant mechanisms of information acquisition and manipulation in autocracies.

To focus our review, we restrict attention to fully specified game-theoretic models, although it bears emphasis that much of the literature traces its intellectual roots to the less formal but seminal work of Geddes (1999), North & Weingast (1989), Olson (1993), Tullock (1987), and Wintrobe (1998). Throughout, we intentionally focus on a small number of key models, which we present with enough detail to highlight the main substantive and methodological challenges that arise, while only briefly outlining the key arguments in the broader, related literature.

**REGIME CHANGE, REVOLUTION, AND REPRESSION**

The overarching desire of all governments is to stay in power. In democracies, the rules of engagement between the government and the opposition are delineated constitutionally. There
are no such guarantees in dictatorships. A dictator may be overthrown nonconstitutionally—by his inner circle, by the military, or by revolution. What then keeps an authoritarian leader in power? How do authoritarian leaders maintain the support of powerful factions without ceding control to them? What are the consequences and limits of repression as a strategy of political survival?

In answering these questions, one difference between democracies and dictatorships is essential: Actors in the latter cannot rely on the commitment power provided by various institutions in democracies, especially independent high courts. In the extreme, no commitments are possible under dictatorship.

Acemoglu et al. (2008) develop a model that clarifies the challenges to regime stability of an inability to make credible commitments. In their model, some members of a ruling coalition may eliminate others if they command greater power than those being eliminated. Critically, however, those who survive cannot credibly commit to refrain from subsequently eliminating other surviving members. This process therefore continues until no coalition can be formed to eliminate further members—perhaps because members of that coalition fear their own subsequent elimination. The model illustrates that stable coalitions can arise even without institutions or a single dominating force (e.g., a leader with the power to punish deviators). At the same time, the ultimate stable coalition is not necessarily minimal (cf. Riker 1962), and the member with the greatest power may be eliminated in the process of forming a stable ruling coalition.

**Model: Dynamic Control of Power**

Consider a society with members \( A, B, C, \ldots \), where each member \( i \) is endowed with power \( y_i \geq 0 \). Suppose each member of society aims to maximize his proportional share of a divisible resource (a “pie”). The distribution of this pie is determined through a process of coalition formation, modeled as follows. Beginning with the initial coalition (i.e., the entire society), any subcoalition can form to expropriate the other members of the coalition, provided that the subcoalition has greater aggregate power than the remaining members. At the conclusion of this process of subcoalition formation and elimination, the pie is divided among the surviving members in proportion to their power. The ultimate ruling coalition must satisfy precisely one requirement to be stable: It must not contain a stable subcoalition whose members unanimously prefer to eliminate at least one member of the ruling coalition and have the power to do so.

To illustrate the insights provided by this model, consider first a two-member society with members \( A \) and \( B \). If \( y_A > y_B \), then beginning with the coalition \( \{ A, B \} \), \( A \) eliminates \( B \) and acquires the entire pie for herself. Conversely, if \( y_A < y_B \), then \( B \) emerges as a one-member stable ruling coalition. Two-member coalitions are therefore generically (i.e., as long as \( y_A \neq y_B \)) unstable.

Now, consider a three-member society \( \{ A, B, C \} \), and suppose \( y_A < y_B < y_C < y_A + y_B \). Given this power distribution, no member can form a subcoalition of her own. Further, no two-member subcoalition will form, given that two-member coalitions are generically unstable. To see this, suppose \( A \) and \( B \) considered eliminating \( C \). They could do so, as \( y_A + y_B > y_C \), but \( A \) would not agree to this because she would subsequently be eliminated by \( B \). The three-member coalition \( \{ A, B, C \} \) is therefore a stable ruling coalition.

Finally, consider a group consisting of four individuals, \( A, B, C, \) and \( D \), with power \( y_A = 3 \), \( y_B = 4 \), \( y_C = 5 \), and \( y_D = 10 \). Although \( D \) is most powerful, his power is not sufficient to eliminate the coalition \( \{ A, B, C \} \). Crucially, however, \( D \) alone is stronger than any two members of \( \{ A, B, C \} \), which implies that any three-member coalition that includes \( D \) is unstable. But \( \{ A, B, C \} \) is
stable: In equilibrium, these three members form a stable ruling coalition that excludes the most powerful individual!\(^4\)

**Discussion**

This framework can be used to examine a variety of questions about authoritarian politics. Acemoglu et al. (2012), for example, study a dynamic model in which current decision makers, in addition to choosing policy, determine the identity of decision makers in the next period (this can be interpreted as the choice of a constitution). In principle, actors may reap policy benefits by transferring decision-making power to another group, as when the rich avoid expropriation by enfranchising the poor (see Acemoglu & Robinson 2000, 2001, 2005). Nonetheless, in the spirit of Acemoglu et al. (2008), decision makers may refrain from adopting a constitution that provides such policy benefits if they expect it to eventually lead to an even less favorable constitution.

A different perspective on coalition formation arises from models of “divide and rule,” in which a leader plays factions of society against each other. Acemoglu et al. (2004) analyze a model in which the ruler threatens to redistribute tax revenue from those who oppose him to those who support him. As this threat can be credibly made to all groups, the ruler is able to remain in power and keep the tax revenue for himself.\(^5\) Meanwhile, Padró i Miguel (2007) posits that the fear that power would pass from one (ethnic or sectarian) group to another prevents citizens from holding rulers from their own group accountable. In a similar spirit, Svolik (2009) develops a model in which the conjunction of imperfect information about the dictator’s actions and commitment problems allows the dictator to usurp power at the expense of his supporters, possibly culminating in the emergence of personal autocracy.

The formation and maintenance of authoritarian ruling coalitions comprise just one challenge that authoritarian leaders face. Opposition from outside the ruling coalition—typically in the form of popular protest and revolution—presents another major challenge to autocratic stability. A number of models examine the dynamics of collective action in antiregime protests and the means by which dictators can prevent their success.

An important intuition in the study of protest and revolutions concerns “cascades”: the possibility that a protest today spurs more protests tomorrow by revealing information about the degree of popular support for the regime. Kuran (1991) attributes such a dynamic to the heterogeneous tolerance for “preference falsification” among the population. Given this heterogeneous, an initial protest by a few dissidents may encourage a wider protest by citizens with a smaller distaste for conformity, which may in turn encourage an even broader protest, and so forth. Lohmann (1993, 1994) models a complementary informational rationale according to which citizens who are only partially informed about the state of the world (e.g., the competence or vulnerability of the regime) update their beliefs based on the protest activity of others. Yet, as Meirowitz & Tucker (2013) show, information revelation via protest may also have the opposite consequence: If a series of “color revolutions” fails to produce a better government, the opposition may conclude that the problem lies not with any specific government but with the society as a whole—a draw of a good government is simply too unlikely.

---

\(^4\) Acemoglu et al. (2008) suggest that this stylized example corresponds to the elimination of Lavrentii Beria (with the power 10) by other members of the Soviet Politburo following Stalin’s death. But note that a coalition with the power distribution \(3, 4, 5, 10, 20\) is stable; think of the last, most powerful member being Stalin.

\(^5\) See Roemer (1985) for an early model of redistribution in the pursuit of nondemocratic power.
Several recent papers adopt a similar informational perspective, modeling revolutionary activity as a coordination game in which citizens receive private signals about some politically relevant fact (e.g., Casper & Tyson 2014, Edmond 2013, Egorov & Sonin 2014, Hollyer et al. 2015, Persson & Tabellini 2009, Shadmeir & Bernhardt 2011). An interesting question that arises in these settings is to what extent elites can influence coordination by the masses. Bueno de Mesquita (2010) shows that revolutionary vanguards can use violence to strengthen the perception that most of the population opposes the regime, yet at the same time, vanguards engage in the most violence precisely when citizens are already likely to successfully coordinate on protest.

What can authoritarian rulers do to prevent regime overthrow? At a general level, this is the subject of the remainder of this review. But most immediately, they can raise the costs of political participation via repression. An extreme version of this strategy is the physical elimination of citizens inclined to oppose the regime (Gregory et al. 2007). The result is a shift of the median citizen’s ideal point closer to the government’s ideal point. In a model of protest that incorporates such spatial considerations, Dagaev et al. (2013) show that the regime can reduce the attractiveness of revolution (and thus increase its probability of survival) by adopting policies close to the median citizen’s ideal point.

Other models emphasize the limits of repression. In a model similar in spirit to Lohmann (1993, 1994), Kricheli et al. (2011) show that repression is a double-edged sword. Repression reduces the risk of protest, but if protest nonetheless occurs, it is more likely to be effective: Citizens who protest in a repressive environment send a strong signal that they oppose the regime. Meanwhile, Acemoglu et al. (2010b) and Svolik (2013a) point out that reliance on the military for repression endows that organization with the capacity to demand concessions from, and potentially overthrow, the government. Similarly, Egorov & Sonin (2011) show that the dictator’s capacity to defend himself against challengers—by repression or otherwise—is limited by the possibility of treachery by his lieutenants. Finally, Lorentzen (2013) demonstrates that authoritarian regimes may in fact benefit from localized, limited protests, as by doing so they gather information that is otherwise hard to collect in the absence of competitive elections or free media.

INSTITUTIONS

Game-theoretic models of institutions in dictatorships address a number of puzzles that are central not only to authoritarian politics but to political science and political economy more broadly: Do institutions have any independent power or are they mere reflections of underlying power relations? How can institutions constrain leaders in political systems where violence is a frequent and often the ultimate arbiter of conflicts? Can institutions alleviate commitment and credibility problems that plague societies where rule of law is weak or nonexistent? How can institutions bind the very same actors who adopt or devise them?

As with formal political theory more broadly, the formal analysis of authoritarian institutions usually adopts one of two analytically distinct approaches. The first approach studies the equilibrium consequences of institutions: It takes a particular institution as given and examines how the incentives created by that institution shape the behavior of its participants. The second approach examines institutions as equilibria, treating the emergence and persistence of political institutions as the equilibrium outcome of strategic interaction.

The selectorate theory of Bueno de Mesquita et al. (2003) is an illustration of the first analytical approach—the study of the equilibrium consequences of authoritarian institutions. According to

---

selectorate theory, any political system can be usefully characterized by the following sets of actors: the general population, the selectorate of size $S$, a winning coalition of size $W$, and the leader. The selectorate is the subset of the population whose members may become part of the leader’s support base, whereas the winning coalition is the subset of the selectorate whose support the current leader needs to stay in power. Other than assuming that $W \leq S$, selectorate theory does not place any restrictions on the size of the winning coalition relative to the selectorate. We can therefore compare political systems ranging from democracies, where $W/S \approx \frac{1}{2}$ and the selectorate comprises all citizens eligible to vote, to authoritarian regimes that rely for survival on a significant fraction of the population or elite, $0 < W/S < \frac{1}{2}$, to personal autocracies, where $W/S \approx 0$.

Bueno de Mesquita et al. (2003) explore the implications of such variation in $W/S$ across political systems—and especially across various types of nondemocracies—for their stability, quality of governance, provision of public versus private goods, and participation in international conflict. As they demonstrate, leaders of systems in which $W/S$ is comparatively small can retain the loyalty of their winning coalition at relatively little cost. To understand this argument, consider the following simplified, one-period version of the selectorate model.

### Model: Selectorate Theory

An incumbent leader survives in power as long as he is supported by a winning coalition of size $W$. A challenger, who would like to replace the leader, must gain the support of an alternative winning coalition (also of size $W$), of which at least one member must be a defector from the leader’s winning coalition. In order to maintain his hold on power, the incumbent leader offers each member of his winning coalition a reward $w_i \geq 0$. We denote the reward that the challenger promises to the members of his alternative winning coalition by $w_c \geq 0$. A key assumption is that the incumbent (or the challenger) must offer the same reward to all members of his winning coalition.

In order to be credible, both rewards must satisfy a budget constraint: For all $j \in \{I, C\}$, $w_j W \leq R$, where $R$ denotes the government’s revenue. But the challenger faces an additional credibility problem: Upon seizing power from the incumbent, he may prefer to replace some members of the winning coalition that brought him to power with others from the selectorate. In particular, assume that if the challenger gains power, he will then form his final winning coalition from the $W$ members of the selectorate for whom he has the highest “affinity.” To keep the analysis simple, suppose that the challenger’s affinity $A_i^C$ for any member $i$ of the selectorate is drawn from the standard uniform distribution, with affinities realized only after the challenger takes power.

A member of the winning coalition who considers defecting to the challenger therefore expects to become a member of the challenger’s ultimate winning coalition only if he is among the $W$ members of the selectorate with the highest realization of the affinity parameter $A_i^C$, which occurs with probability $W/S$. [For a proof, see Supplemental Material (follow link from the Annual Reviews home page at http://www.annualreviews.org)]. By contrast, members of the incumbent’s winning coalition know that they are among the $W$ members of the selectorate for whom the incumbent has the highest affinity, as they are already in the winning coalition.

How is authoritarian governance shaped by the relative sizes of the winning coalition and selectorate? A member of the selectorate who is not a member of the incumbent’s winning coalition is

---

7 For a model of authoritarian accountability that builds on selectorate theory, see Besley & Kudamatsu (2008).

8 For a more comprehensive but parsimonious presentation of selectorate theory that incorporates spending on public goods, see Gehlbach (2013, pp. 143–52).
happy to join the challenger for any \( w_C \geq 0 \). But any potential defector from the incumbent’s winning coalition must be compensated for her uncertainty about being included in the challenger’s final winning coalition once the latter unseats the incumbent. More specifically, a member of the incumbent’s winning coalition is willing to defect only if \( \frac{w_I}{w} \geq w_C \), or equivalently, only if the challenger offers at least \( \frac{w_I}{w} \) times as much as the incumbent. The incumbent knows that the largest reward the challenger can promise in light of the budget constraint is \( w_C = \frac{R}{W} \). In turn, if the incumbent wants to preclude any defections, he must offer each member of the winning coalition a reward \( w_I \geq \frac{R}{W} \cdot W_S = \frac{R}{S} \). Put differently, a leader who pays each member of his winning coalition \( w_I = \frac{R}{S} \) is safe from defections and keeps for himself \( R - W \cdot \frac{R}{S} = R(1 - \frac{W}{S}) \) of the country’s revenue.

**Discussion**

Selectorate theory allows us to translate prominent institutional features of nondemocracies—including whether they have legislatures, single parties, military juntas, or monarchical councils—onto the two-dimensional space defined by the sizes of the winning coalition and the selectorate, and to examine the implications of those institutional features.\(^9\) The above model, for instance, implies that leaders of regimes with a winning coalition that is small relative to the size of the selectorate (i.e., small \( \frac{W}{S} \)) will be able to maintain the coalition members’ “loyalty” at a smaller price. Selectorate theory thus helps us to explain why, as nondemocratic regimes become more personalist, their leaders become more repressive and kleptocratic not only toward the general population but also toward their inner circle—and yet, nonetheless, survive in office.

Conceptually, selectorate theory arrives at intuitive but far from obvious implications about leader survival by taking differences in \( W \) and \( S \) across regimes as given and examining their equilibrium implications. But selectorate theory is just one example of the general analytical approach that takes authoritarian institutions as given and examines the implications of institutional differences across dictatorships for their survival and policies [see especially the pioneering work of Geddes (1999) and Wright (2008)]. This approach raises a key question in the study of authoritarian institutions: Why and how are dictatorships able to maintain politically consequential institutions in the first place?

This question is explicitly addressed by the second approach to formal analysis of institutions in dictatorships, which requires that political behavior and the institutions that govern it jointly constitute an equilibrium: Institutions themselves must be equilibria. This requirement is key to answering another major puzzle for formal as well as nonformal theories of authoritarian politics: How do institutions contribute to the stability of dictatorships?

A prominent intuition developed by a number of models is that authoritarian institutions may serve to alleviate commitment problems, which are pervasive in authoritarian politics. As we emphasized above, dictatorships inherently lack an independent, third-party authority with the power to enforce agreements among key political actors, including the leader, economic and

---

\(^9\) In this simplified model, we are somewhat inconsistent in assuming that the challenger can commit to reward members of his winning coalition by a particular amount \( w_C \) but cannot commit to whom the reward will go. The inconsistency disappears in an infinite-horizon setting, in which members of the challenger’s “transition coalition” may or may not ultimately become members of the winning coalition formed after the challenger seizes power; see Bueno de Mesquita et al. (2003) and Gehlbach (2013, pp. 141–52).

\(^{10}\) The claim that diverse political systems can be effectively characterized by a small number of parameters is reminiscent of veto-players theory (Tsebelis 2002).
political elites, repressive agents, and the masses. Thus, as in our simple model of selectorate theory, a challenger may promise to share the spoils of office with those who will help him overthrow the incumbent but face the temptation to renege on that promise once he succeeds. But this is only one among many commitment problems that pervade authoritarian politics: An incumbent authoritarian leader may agree to share power with elites in exchange for their support against a challenger but renege on that promise once his hold on power is secure (Boix & Svolik 2013, Myerson 2008); the regime may encourage investment by economic elites but later prey on the fruits of that investment via confiscatory taxation or outright expropriation (Gehlbach & Keefer 2011); the leader may promise organizational privileges and resources to the military in exchange for repressing a mass protest or insurgency but withdraw them once that threat is eliminated (Acemoglu et al. 2010b, Svolik 2013a); the elite may redistribute wealth to the masses during a revolutionary moment but renounce such concessions once the revolutionary threat subsides (Acemoglu & Robinson 2001, 2005; Boix 2003).

In each of these scenarios, the leader’s promise may lack credibility because, once various actors have expended costly effort to make him more secure (ranging from explicit support to foregoing the option to revolt), the leader will be tempted to renege on his commitment to reward them—and there is no independent authority with the power to stop him. Anticipating this commitment problem, these actors will be reluctant to aid the leader in the first place, thus undermining the regime’s stability.

Commitment problems are thus at the heart of authoritarian survival. A dictator needs to be able to credibly commit to rewarding his supporters, whether it is for helping him to acquire power or for defending him against threats to his hold on power. Myerson (2008) illustrates how institutions may resolve such commitment problems. Rather than assuming that institutions have independent power to enforce the leader’s promises, institutions in Myerson’s model enable the leader’s supporters to coordinate on abandoning him, should he renege on his commitment. Without the institution, notables worry that the leader could betray some of them and still count on the others’ support; with the institution, the leader’s defection becomes common knowledge. Perhaps paradoxically, dictators can improve their chances of survival by establishing institutions that make them vulnerable to their supporters.

Model: The Dictator’s Commitment Problem

Assume a polity consisting of an incumbent leader and two notables, $i = 1, 2$. In each of two periods, the leader faces a challenge to his rule that can only be met with the assistance of the notables. In particular, at the start of each period $t$, each notable $i$ chooses an effort level $e_{it} \in \{0, 1\}$. We say that notable $i$ exerts effort in period $t$ if $e_{it} = 1$. The leader survives (to fight another day, if $t = 1$) with probability 1 if both notables exert effort, with probability $\lambda \in (0, 1)$ if only one does, and with probability zero if neither does. The notables’ fate is tied to the incumbent’s: If the incumbent falls, the flow of payoffs for all actors drops to zero.

Each notable bears a cost $c > 0$ for each period in which she exerts effort. In period 1, if the leader survives, he may (partially) compensate this cost by choosing a wage $w_i \geq 0$ to provide

---

11In principle, questions about the existence of such an independent authority may also be raised in democracies. High-court rulings or legislative acts may, after all, be ignored or disobeyed. But such outcomes would by definition amount to the breakdown of a democratic constitutional order.

12We treat the challenger as a nonstrategic actor whose capacity to replace the leader is fixed. The parameter $\lambda$ may be interpreted as the leader’s power, quality, or wealth, or as some structurally determined balance of power between incumbents and challengers.
to each notable $i$. Critically, we do not assume that the leader can credibly commit to this wage ex ante. In equilibrium, it must be in the leader’s interest to provide this wage after the notables have decided whether to exert effort. Further, each notable receives an exogenous payoff $r > 0$ if the leader survives through the second period. The leader similarly receives an exogenous payoff $R > 0$. As is typical in such models, the payoffs $r$ and $R$ can be interpreted as continuation values from an infinite-horizon game.

To focus on the interesting case, we assume

$$\lambda r < c < r < 2c. \quad 1.$$ 

Observe that the first two inequalities imply that the effort game in period 2 has two pure-strategy Nash equilibria: one where both notables exert effort and one where neither does. The key to whether it is possible to induce effort in period 1 is whether notables can coordinate on one of these equilibria, given the leader’s payment of wages. To examine this question, we explore two versions of the model. In the first version, we assume the presence of an institution, such as a legislature, the key function of which is to allow each notable to observe whether the leader has paid not only her but also the other notable. In the second version, each notable observes her own wage payment but not that of the other notable.

Consider first the version of the model with institutions. Given perfect information, the notables can condition their joint effort in the second period on whether the leader paid each notable i a wage $w_i \geq \bar{w}$, where $\bar{w}$ is the minimum wage that justifies effort across both periods. Jointly, the cost of effort $c$, the exogenous payoff from surviving through the second period $r$, and the fact that the leader (and notables) survives with certainty if both notables exert effort in each period imply that $\bar{w} = 2c - r$. It is optimal for the leader to pay this wage to both notables if the expected payoff from doing so is greater than that from deviating to paying only one notable,

$$R - 2\bar{w} \geq 0. \quad 2.$$ 

In Inequality 2, the right-hand side of the inequality follows from the notables’ ability to coordinate on no effort if the leader fails to provide the promised wage to either. (The expected payoff from providing the wage to neither notable is also zero.) Using the value of $\bar{w}$ derived above, Inequality 2 can be rewritten as

$$R + 2r \geq 4c,$$

which can be interpreted as an efficiency condition: The total benefits from autocratic survival are greater than the costs of ensuring that outcome.

Now consider the case of no institutions. Because each notable observes only whether she has been paid for her effort, the notables cannot coordinate on no effort if only one of them has not received $\bar{w}$. Failure to pay $\bar{w}$ to only one of the notables thus results in only that notable withdrawing effort in the second period. This consequence will discourage the leader from reneging on his promise to pay $\bar{w}$ to both notables if, in addition to Inequality 2, the expected payoff from paying two notables and surviving with certainty is greater than that from paying one notable and surviving with probability $\lambda$,

$$R - 2\bar{w} \geq \lambda R - 2\bar{w},$$

13The first and second inequality imply, respectively, that at the beginning of period 2, a notable is not willing to exert effort if he is the only one doing so, but is willing to exert effort if the other notable does so as well. The last inequality implies that a positive wage must be paid to each notable to motivate effort across both periods.

14To support this behavior, any notable who does not receive $\bar{w}$ must believe, with sufficiently high probability, that the other notable also did not receive $\bar{w}$. 

574 Gehlbach • Sonin • Svolik
or equivalently, when

\[ \lambda \leq \frac{R - (2c - r)}{R}. \]

Incentivizing effort that will ensure the leader’s survival is thus easier with than without institutions, as in the latter case, an equilibrium in which both notables exert effort exists for a smaller region of the parameter space.

But not all dictatorships benefit from establishing institutions. In this model, it is leaders with intermediate power, that is, those for whom

\[ \frac{R - (2c - r)}{R} < \lambda < \frac{c}{r}, \]

who face a commitment problem that institutions can solve.\textsuperscript{15} By contrast, when \( \lambda \leq \frac{R - (2c - r)}{R} \), the leader is so weak that he can credibly commit to pay both notables, even without institutions. Meanwhile, when \( \lambda \geq \frac{c}{r} \), the leader is so strong that each notable strictly prefers to exert effort in the second period, regardless of what the leader has done or what she expects the other notable to do.\textsuperscript{16}

Discussion

Institutions play a conceptually similar role in the model of Gehlbach & Keefer (2011), where an authoritarian party helps its members to learn about the expropriation of others by the leader; in the model of Boix & Svolik (2013), where legislatures allow the dictator’s supporters to monitor his compliance with their agreement to share the spoils of office; and in the model of Magaloni (2008), where the formal rules of parties and elections, via which the dictator delegates key powers, help his allies to monitor the dictator’s compliance. Meanwhile, a related class of models emphasizes the role of authoritarian institutions in facilitating cooptation, either by providing a forum in which bargaining over concessions can proceed effectively (Gandhi 2008, Gandhi & Przeworski 2006) or by structuring intertemporal incentives that encourage sunk political investments by party members and thus generate greater loyalty to the regime (Lazarev 2005; Svolik 2012, ch. 6).

Across these models, institutions accomplish their political functions—such as credible commitment or cooptation—by reducing asymmetries of information among their participants. This occurs either because participation within institutions implies better opportunities for communication or because the existence of formal rules of interaction makes their violations easier to observe. These mechanisms thus help us to answer a more general question: What is it about formal institutions that allows them to accomplish politically beneficial outcomes that other, non-institutional forms of interaction cannot?\textsuperscript{2}

ELECTIONS AND MEDIA CONTROL

The model of the dictator’s commitment problem in the previous section emphasizes the role of institutions in providing information about the dictator’s actions. This is just one of the many types of information that are dispersed among elites and citizens in dictatorships. In some situations, autocrats may benefit from acquiring such information or from making it broadly available. In others, governments may prefer to prevent particular information from becoming public or they

\textsuperscript{15} The second inequality follows from Condition 1.

\textsuperscript{16} This scenario, depending on the value of \( \lambda \), may result in additional commitment problems and a greater role for institutions.
may manipulate information flows so that political and economic actors take actions that best serve the regimes’ objectives. The models of elections and media control that we review in this section address these possibilities.

Elections are either mere facades or altogether absent in some dictatorships, as in the cases of Nazi Germany, the Soviet Union, and Saddam Hussein’s Iraq. Yet many nondemocracies do have elections that are at least somewhat competitive—Mexico under the Institutional Revolutionary Party (PRI) being a prime example (Magaloni 2006)—and such “electoral authoritarianism” has been on the rise since the end of the Cold War (e.g., Levitsky & Way 2010). Various models explain the presence of elections and the role of electoral manipulation in autocracies.

A key question addressed by this research agenda is why nondemocratic regimes would allow competitive elections in the first place. Little (2014) usefully distinguishes between models in which the regime uses elections to gather information about its popular support versus models in which elections signal the regime’s hold on power to external audiences. A seminal contribution in the former tradition is Cox’s (2009) model of a bargaining game between an autocrat and opposition. In the spirit of Fearon (1995), information asymmetries create the possibility of a bargaining failure in the form of a coup or revolution. The role of elections is to provide the autocrat with information about his regime’s popularity, which he uses to infer an appropriate level of concessions. Consistent with this theoretical perspective, Cox (2009) finds that multi-party elections discourage violent turnover in autocracies.

A second, distinct class of models emphasizes the ability of elections to signal the regime’s strength to various audiences. Egorov & Sonin (2014), for example, model an autocrat who has private information about his popularity (e.g., because of internal polling) and chooses from a menu of options to secure his hold on power. Popular dictators prefer to hold competitive elections, as these discourage rebellion by revealing the regime’s popularity. In a similar spirit, Little (2012) shows that incumbents are more likely to hold competitive elections when their potential impact on citizens’ beliefs is large. In Little’s model, this occurs when the regime’s popularity is neither too high nor too low.

Thinking about elections as mechanisms for signaling strength suggests a trade-off. A competitive election allows an autocrat to more credibly communicate his popularity to various audiences, yet this entails the risk of an unfavorable election outcome (either an outright loss or disappointingly narrow victory). Various models consider the role of electoral manipulation in managing this trade-off. The basic idea is that fraud can be effective in manipulating beliefs even when it is commonly known that fraud has taken place. To see this, consider the following simple framework, adapted from Rozenas (2012, 2014b) and Gehlbach & Simpser (2015).

**Model: Electoral Manipulation**

There is an incumbent ruler of unknown popularity—for simplicity, \( \theta \in \{0, 1\} \)—and an opposition. The incumbent and opposition share a common prior belief that the incumbent is popular (i.e., that \( \theta = 1 \), which can be interpreted as meaning that the incumbent has the support of a majority of the population) with probability \( \gamma \). At the beginning of the game, the incumbent decides whether to manipulate an election, \( m \in \{0, 1\} \), the outcome of which, \( v \in \{0, 1\} \), is jointly determined by the incumbent’s popularity and manipulation decision. In particular, if the incumbent is popular, he wins the election (\( v = 1 \)) with certainty, whereas if the incumbent is unpopular, he wins the election with probability \( m \phi \), where the parameter \( \phi \in (0, 1) \) measures the effectiveness

---

17For a model of endogenous elections in which there is no uncertainty about either variable, see Przeworski et al. (2012).
of manipulation. The opposition observes the manipulation decision and the election outcome (but not the incumbent’s popularity), following which it decides whether to challenge the incumbent. A challenge is successful if the incumbent is unpopular and unsuccessful if he is popular. Normalize the payoff for both incumbent and opposition from holding/seizing power to one, and let the cost of manipulation and of challenging the incumbent be $\kappa$ and $\zeta$, respectively, where $\kappa > 0$ and $\zeta \in (0, 1)$.

The incumbent’s manipulation decision depends on the expected response of the opposition. With no manipulation, the opposition fully updates its beliefs about the incumbent’s popularity based on the election result, challenging if and only if $v = 0$. In contrast, if the incumbent manipulates, by Bayes’ Rule the opposition believes the incumbent to be unpopular with probability $\frac{(1 - \gamma)\phi}{\gamma + (1 - \gamma)\phi}$ when $v = 1$. A necessary condition for manipulation is therefore

$$\zeta \geq \frac{(1 - \gamma)\phi}{\gamma + (1 - \gamma)\phi};$$

else the opposition would challenge regardless of the election result. Assuming this condition holds, the incumbent’s choice reduces to a trade-off between the cost of manipulation $\kappa$ and the greater probability of obtaining a favorable election result. Thus, the incumbent manipulates if

$$\gamma + (1 - \gamma)\phi - \kappa > \gamma.$$  

Combining Conditions 4 and 5, manipulation occurs in equilibrium if

$$\frac{\kappa}{1 - \gamma} < \phi < \frac{\gamma\zeta}{(1 - \gamma)(1 - \zeta)}.$$  

### Discussion

That electoral manipulation can be effective even when it is observed—as in this example—is a consequence of the assumption that the mapping from fraud to election outcomes is uncertain. One cannot perfectly infer the incumbent’s popularity from the election result, even if it is commonly known that some manipulation occurred. One motivation for this assumption is that fraud is carried out in a decentralized fashion by the regime’s agents: It is easy enough to observe that fraud has taken place but difficult to know its true extent. An additional implication of this perspective is that manipulation may be more or less effective than is optimal from the incumbent’s perspective (in the setting above, Condition 6 may not be satisfied). As Rundlett & Svolik (2015) show, when rewards or sanctions for engaging in manipulation are contingent on the incumbent’s survival [a key assumption of Gehlbach & Simpser (2015) as well], it will be rational for the regime’s agents to engage in manipulation only if they expect other agents to do so. This strategic complementarity implies that semicomp etitive elections may be won or lost by unnecessarily large margins. In a similar fashion, as Simpser (2013) argues, manipulation that takes the form of threats or promises to encourage turnout can result in herd behavior among voters, again producing abnormally large margins of victory.

A related question about authoritarian elections is why autocrats sometimes tolerate the presence of independent election monitors, electoral commissions, and other third parties who could undermine electoral fraud. The model above suggests an explanation: In the absence of such actors, manipulation may be “too effective” from the incumbent’s perspective. In particular, greater electoral transparency can reduce the risk of violent postelection conflict, as when monitors encourage coordination on formal rules (Fearon 2011, Little et al. 2014) or certify a narrow incumbent victory that might otherwise be interpreted as fraudulent (Magaloni 2010, Chernykh & Svolik 2015). Moreover, when citizens can use the level of manipulation to divine the “true” election result, fraud
may be inefficient from the incumbent’s perspective; allowing for election monitors provides the incumbent with a way to tie his hands (Little 2012, 2015). Nonetheless, when the incumbent can choose from a menu of manipulation strategies, greater transparency may encourage the use of less observable forms of fraud, such as when incumbents engage in “election night” manipulation, changing polling results after they have come in (Rozenas 2014a).

Electoral institutions and electoral fraud thus play an important role in regulating information flows in autocracies. The media—as a primary source of information for many citizens—are similarly critical. Indeed, although violence has historically been an important instrument of authoritarian governance, modern dictators often rule by “velvet fist,” relying on manipulation of the media and other sources of information to remain in power and pursue policy goals (Guriev & Treisman 2015). From a theoretical perspective, this strategy raises two questions. First, why is censorship effective, given that citizens are usually aware of a progovernment bias in news reports? Second, from the dictator’s perspective, what are the trade-offs involved in control of the media?

We address these questions by focusing on two related classes of models: those that emphasize media control as a means of persuasion and those that focus on the role of censorship in demobilizing a dissatisfied public. Gehlbach & Sonin (2014) develop a model of the first mechanism, in which media control allows the ruler to persuade citizens to take actions that may not be in their individual best interest. A key conclusion is that media control can be effective, even when citizens are aware that news reports are biased. Nonetheless, bias comes at a cost to the ruler.

Model: Media Control

Suppose a state of the world can take two values, $s \in \{0, 1\}$, where any citizen prefers to take some action desired by the ruler if and only if $s = 1$. Citizens and ruler share a common prior belief that $s = 1$ with probability $\theta$. Prior to realization of the state $s$, the ruler chooses an editorial policy $\beta$, which is the probability that a media outlet reports that $s = 1$ when in fact $s = 0$.\(^{(18)}\) (When the media outlet is privately owned, the ruler induces this editorial policy through threats or subsidies.) After observing $\beta$, citizens decide whether to watch the news, which entails an idiosyncratic cost; only those who watch the news receive the report. Finally, after the media outlet delivers its news report—determined by the state of the world and the editorial policy—citizens decide whether to take the action desired by the ruler.

By assumption, citizens observe the editorial policy—that is, media bias—chosen by the ruler. Nonetheless, media control can be effective if the media mix enough fact with fiction to keep citizens guessing. [This is a special case of Bayesian persuasion, as discussed by Kamenica & Gentzkow (2011).] To see this, observe that, by Bayes’ Rule, citizens’ posterior belief that the state $s = 1$, conditional on having observed a media report to that effect, is $\theta \frac{\beta}{\theta \beta + (1 - \theta)(1 - \beta)}$, which is greater than the prior belief $\theta$ if $\beta < 1$. More generally, the degree to which citizens are persuaded by a positive media report is negatively related to the degree of media bias (that is, to $\beta$).

From the ruler’s perspective, increasing the level of media bias $\beta$ thus involves a trade-off. On the one hand, conditional on watching the news, citizens are more likely to receive a report that $s = 1$ when $\beta$ is large. On the other hand, they are less likely to believe such a report, which discourages them not only from taking the action desired by the ruler, but also from watching the news to begin with.\(^{(19)}\)

\(^{(18)}\)It is always optimal for the ruler to choose an editorial policy that truthfully reports the state when $s = 1$.

\(^{(19)}\)A similar trade-off motivates Shadmehr & Bernhardt’s (2015) argument that an autocrat might wish to precommit to a censorship level, as assumed here.
Discussion

Various political and economic factors influence the trade-off identified in this model of media control. One important variable is the size of the advertising market (Besley & Prat 2006, Petrova 2012). The ruler chooses a less biased editorial policy when the advertising market is large, as the opportunity cost of lost viewership—incurred directly when the media outlet is state-owned and indirectly through government subsidies when it is private—is greater. At the same time, inducing a private firm to report biased news when the advertising market is large may be so costly that the ruler prefers to nationalize the media instead (Gehlbach & Sonin 2014). Such effects may be particularly pronounced when there are numerous media outlets, as any outlet must be compensated for the full loss of advertising revenue (not only its share), given that it could capture all viewers by deviating from the common bias chosen by the ruler (Besley & Prat 2006).

A related class of models examines the role of media control in demobilizing citizens. The question of how such manipulation could be effective is addressed by Edmond (2013). A key assumption in his model is that the autocrat possesses private information about his vulnerability to mass protest, so that citizens are unable to infer the autocrat’s type from beliefs about the autocrat’s strategy and the signals they receive. This effect is compounded by the strategic complementarity that characterizes mass protests (here modeled as a global game): since most citizens prefer to participate only if they expect others to do so as well, a small amount of manipulation can have a large impact on mobilization.

Although insightful about the reasons that media control can be effective, Edmond (2013) simply assumes that manipulation is costly to the autocrat. Other models endogenize this cost and so consider the trade-offs between demobilization and other considerations. Egorov et al. (2009) show that free media provide autocrats with information that can be used to incentivize bureaucratic effort, even as it raises the risk of a coordinated uprising in the event of poor performance. The former consideration dominates the latter when bureaucratic performance is especially important—for example, when resource rents are few. Lorentzen (2014) similarly considers the trade-off between encouraging investigative reporting in order to curb corruption by local officials and the risk that such reporting could result in a mass uprising by exposing large discontent at the national level. Finally, Chen & Xu (2014) consider how horizontal information flows among citizens may either encourage or discourage collective action, depending on whether citizens agree with each other about government policies.

Conclusions

Roger Myerson wrote in 2008, “There are relatively few modern game-theoretic models of autocratic politics, and more are needed” (Myerson 2008, p. 125). Since that article was published, scholars have heeded its call and provided many formal models of nondemocratic politics, as this review has demonstrated. Nonetheless, the field is in its infancy. Paradoxically, the oldest and most common form of political governance is the least studied with tools of modern social science.

In writing this review, we hoped to organize the discussion around a single general model that would capture the range of insights developed by the many particular models and theories of nondemocratic politics. This proved impossible, for reasons that we believe are illustrative of the field. Unlike the study of democratic politics, which is substantially organized around the spatial

Edmond (2013) draws a contrast to career-concerns models à la Holmström (1999), in which the absence of private information implies that citizens have correct beliefs in equilibrium not only about the sender’s strategy but also about his particular action.
model—Brady (2011) refers to it as political science’s “brand”—there is no canonical model of authoritarian politics. Rather, the field consists of a grab-bag of models that exploit the various techniques and practices of modern game theory.21

Such small-t theorizing—not a pejorative, in our view—is an understandable and possibly necessary consequence of the conventional understanding of dictatorship as a residual category, encompassing any regime that fails to meet established criteria for democracy. Dictatorships are in turn associated with an extraordinary diversity of institutions, leaders, policies, and outcomes—unified primarily by the fact that their political system is not democratic. Accordingly, some of the models that we reviewed highlight quintessentially authoritarian mechanisms (e.g., repression and violent leadership change), while many others address particular departures from democratic practice (e.g., electoral manipulation and media control) or the distinctive role in autocracies of superficially democratic institutions (e.g., legislatures and elections).

Notwithstanding the diversity of the field, the study of nondemocratic institutions does share some broad insights and mechanisms. Models of electoral authoritarianism and manipulation, for example, largely assume that elections provide information to some audience but that they do not directly determine the autocrat’s survival. This informational perspective seems natural in the context of nondemocratic politics, where asymmetries of information between key actors and uncertainty about the political environment are paramount. The related view that commitment problems abound—that one cannot simply assume, for example, that authoritarian leaders will cede power upon losing an election—arises because formal institutions have only tentative binding power, with violence always available as a means of resolving political conflicts.

Looking to future work, we suggest that these two themes—asymmetries of information and commitment problems—serve as organizing principles for the field. Autocratic institutions exist largely to aggregate information (typically necessary for the autocrat’s survival), not preferences (as through majoritarian elections). They further serve to allow key actors to make credible commitments. Similar arguments have been applied to the formal analysis of democratic institutions (e.g., Gilligan & Krehbiel 1987), but concerns about asymmetries of information and the credibility of commitments are disproportionately more severe in autocracies.

What is largely missing from the theoretical literature is a comparative analysis of institutions. At present, most work in the literature either shows how autocracies function in a stylized institution-free environment, or demonstrates that autocracies can function better (from the perspective of key actors) in the presence of various institutions.22 But if one accepts the premise that institutions evolve to minimize transaction costs (North 1990), which include incomplete information and limited commitment, then one must ask which institutions work best—not just show that some institutions are better than none.

This is the research frontier in the formal theory of nondemocratic politics: the rigorous comparative study of authoritarian institutions, with the goal of understanding why some institutions are more prevalent than others at certain times and places. Such work will naturally dovetail with the literature on institutional change and democratization, which we have deliberately under-emphasized in this review. It will also build on and inform a burgeoning empirical literature on authoritarian politics, which is worthy of a review of its own.

21In this respect, the modeling of authoritarian politics resembles the theory of industrial organization (see Tirole 1988).

22The focus on particular institutions, rather than comparative institutions, may be a consequence of this functionalism, implicit in the analysis of institutions as equilibria: An institution is shown to be “optimal” in some sense.
DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED


