

“I Alone Can Fix It:” the Strongman Narrative and Leader Support*

Carlo M. Horz[†] Korhan Kocak[‡]

Abstract

Politicians want to maintain public support—especially in the context of democratic backsliding. One widely used instrument is the Strongman Narrative (*SN*): a narrative claiming that a strong leader can improve economic performance. Modeling beliefs about causal processes using directed acyclic graphs (DAGs), we study the conditions under which the *SN* shapes citizens’ support decisions. We find a form of “authoritarian legacy:” past support for authoritarian leaders increases the citizen’s incentives to support the leader today. Moreover, believing in the *SN* increases moderates’ support for the leader. We show that the *SN* is the most desirable narrative for the leader among all three-variable DAGs featuring support, leader strength, and economic performance: it ensures the leader enjoys citizens’ support under the broadest set of conditions. We extend the model to investigate the consequences of citizens caring about democratic backsliding and find that, unless citizens’ costs of backsliding are sufficiently high, the *SN* remains the narrative that is the most desirable for the leader.

*Previous versions were circulated under the titles “How To Keep Citizens Disengaged: Propaganda and Causal Misperceptions” and ““I Alone Can Fix It:” the Strongman Narrative and Democratic Backsliding.” We thank two anonymous reviewers and the editors for detailed feedback. We also like to thank Quintin Beazer, Afiq bin Oslan, Federica Izzo, Holger Kern, Andrew Little, John Patty, Adam Przeworski, Jessica Sun, Ian Turner, Scott Tyson, and audience participants at APSA 2021, MPSA 2024, the Behavioral Models of Politics 2024, the Political Economy of Authoritarianism and Democratic Backsliding conference at Yale, and the Political Economy of Democratic Backsliding conference at Columbia for helpful comments and suggestions on previous versions of the paper. Kocak gratefully acknowledges financial support from the Center of Behavioral Institutional Design and Tamkeen under the NYU Abu Dhabi Research Institute Award CG005.

[†]Assistant Professor, Department of Political Science, Washington University in St. Louis, St. Louis, MO 63130. E-mail: horz@wustl.edu. Web: www.carlohorz.com.

[‡]Assistant Professor, SPEGA, IE University, IE Tower — Paseo de la Castellana 259E, Madrid, Spain. E-mail: kkocak@faculty.ie.edu. Web: www.korhankocak.com.

Politicians aim to maintain public support, even when survival in office is not immediately at stake: policies can be passed more easily, political opponents are less eager to challenge, intra-party rivalries can be shut down, and so on. This is especially true when politicians aim for institutional changes, such as the dismantling of democratic institutions. From Vladimir Putin to Recep Tayyip Erdogan to Victor Orban, the rise of strongmen who muscle their way from electoral victory to a more authoritarian ruling style has been the primary threat to democracies since the end of the Cold War (Frantz, 2018; Levitsky and Ziblatt, 2018). A persistent challenge in political science has been to understand why such moves sustain popular support. Why have Putin, Erdogan, and Orban continued to win elections during their rule, despite their agitations against democratic norms? A key ingredient to explain such support is how citizens perceive the benefits and costs of autocratic rule.¹ In this paper, we study the implications of believing in a narrative that links autocrats' rule to a glorified era of national strength—both politically and economically—under ostensibly benevolent yet decisive rulers, such as the Czars, Sultans, and Pharaohs (Elçi, 2022). We term this causal theory the *Strongman Narrative* (*SN*): a strong leader makes a strong economy more likely. According to this narrative, strong leadership is not only associated with historical national greatness but also a pathway to future economic prosperity. Under which conditions is the *SN* effective—and more effective than alternatives—at strengthening citizens' support for the leader?

To answer these questions, we model the *SN* as a causal graph, following Spiegler's (2016) innovation to analyze perceptions using directed acyclic graphs (DAGs). Our baseline model consists of three ingredients: a representative citizen who chooses whether to support a leader, a historical database, and a DAG that the citizen uses to make sense of this database. The database contains information about past choices and outcomes—what we refer to as the polity's history—including instances of past support for the leader. In the benchmark, the citizen has two concerns. First, she cares about ideology and wants to support a politician who is ideologically aligned. Second, she wishes to

¹Acemoglu et al. (2024) show that experimental interventions that increased Turkish voters' awareness on the social and economic costs of authoritarianism decreased Erdogan's vote share.

obtain high economic performance. Our first main causal graph assumes that economic performance is exogenous, and leader strength is influenced both by the citizen’s support decision and by economic performance. This means that, as in canonical theories of authoritarian politics (e.g., Bueno De Mesquita et al., 2005), a strong economy causes leader strength—because would-be challengers can be bought off—and not the other way around. We call this data generating process the buy-off DAG (henceforth *BO*; see the upper panel in Figure 1). Our second main causal graph is the *SN* (shown in the lower part of Figure 1). The *SN* asserts that citizens’ support causes a strong leader, which in turn causes high economic performance. Hence, like the *BO*, the *SN* serves as a causal guide to make sense of the polity’s history: it influences the citizen’s beliefs about the economic consequences of supporting the leader.

Our baseline model establishes that even when the actual data generating process is the *BO*, the *SN* is the most desirable narrative for the leader, in the sense that it ensures citizens’ support under broadest conditions. Relative to rational expectations, the *SN* increases the citizens’ support for the leader by swaying ideological moderates. How much the *SN* can help the leader depends on the context. In particular, we demonstrate a form of “authoritarian legacy:”² the citizen today is more likely to support the leader when historical leaders enjoyed support in the past.

The intuition is as follows. If the citizen believes in the *BO* DAG, she only supports a leader whose net ideological benefit is positive. Thus, the leader’s support consists of his ideological base. The *SN* allows the leader to expand his support by asserting a positive causal link between support and economic performance that goes through leader strength. Specifically, the *SN* tells the citizen that in order to find the expectation of economic performance given support, one must condition on leader strength. If the leader is indeed strong, the economy must be good. This is because, when *BO* is the true data generating process, a good economy is a necessary condition for leader strength.³

²Here, we adopt the definition of a legacy as “a durable causal relationship between past institutions and policies on subsequent practices or beliefs.” (Beissinger and Kotkin, 2014, p. 7). See Pop-Eleches and Tucker (2017, Ch. 9) for empirical evidence on the relationship between communist legacy and contemporary attitudes towards democracy.

³This is only true for a particular parametrization of the *BO* that we employ in our benchmark analysis. In general, the probability of high economic performance conditional on a strong leader is

But if the leader is weak, the economy can be good or bad. We show that, conditional on a weak leader, the expectation of good economic performance is decreasing in the historical frequency of support (Gurr, 1968, 1970). This is because a higher frequency of past support increases the historical correlation between leader strength and a good economy, which the citizen interprets as an effect of the former on the latter.

This finding implies a *historical complementarity*: past support makes the citizen more likely to support today. This important effect of history is reminiscent of Weber’s conception of *traditional* justification of legitimacy, “sanctified through the unimaginably ancient recognition and habitual orientation to conform” (Weber, 2004, p. 34). Similar to strategic complementarity in standard collective action models, this form of complementarity also breeds equilibrium multiplicity. In particular, we solve for a “Personal Equilibrium,” which is a steady-state probability of support. Here, the current probability of support must be consistent with its historical frequency. We show that for a range of ideological benefits, the model features multiple steady states that differ in the amount of support. This finding explains why aspiring autocrats in countries that look similar in terms of their material circumstances enjoy different degrees of support.

The success of the *SN* at eliciting support depends not only on equilibrium selection, but also on the exogenous probability of high economic performance. There are two relevant ideological benefit thresholds that dictate the citizen’s behavior: the *necessity* threshold, where the citizen *may* support; and the *sufficiency* threshold, where the citizen *must* support.⁴ The necessity threshold decreases monotonically as the probability of good economic performance goes up. This is because a higher probability of high economic performance strengthens the inferred positive effect of support on leader strength. This effect also exists on the sufficiency threshold. But here, there is another effect: a higher probability of good economic performance means that it is more likely to be observed even when the leader is weak, making support redundant. Thus, the threshold that *ensures* support is non-monotonic in the exogenous probability of high economic performance.

larger than the corresponding probability conditional on a weak leader.

⁴Put differently, the necessity threshold is defined as the point such that if ideological benefits are below it, the citizen *never* supports. Similarly, the sufficiency threshold is the point above which the citizen *always* supports.

We show that the *SN* is able to elicit support for some levels of ideological benefits for which the *BO* cannot. This feature makes it desirable for the leader. Next, we ask whether there are other narratives that are more desirable than the *SN* when *BO* is the true data generating process. This is not the case. To show this, we first extend our model to allow strong leadership to emerge under a variety of circumstances, i.e., even when the economy is weak or the citizen does not support. We then demonstrate that the *SN* is the most desirable for the leader among all three-variable DAGs, featuring support, leader strength, and economic performance. In particular, the *SN* is more desirable than a narrative that portrays leader strength as exogenous while arguing that both citizen support and leader strength have a direct (positive) effect on economic performance. This narrative of “working together” induces a difference-in-differences estimator in the citizen’s decision making. Perhaps counter-intuitively, this narrative shrinks the leader’s support base relative to the *BO*.

We then extend the setup and assume that the citizen cares about additional outcomes that are a *consequence* of leader strength. Two possibilities stand out. First, the consequence may represent the compliance decision of another non-strategic group of citizens who comply when the leader is strong. The citizen may care about behaving similarly to other citizens—indeed, previous scholarship shows that in authoritarian regimes, citizens often want to act similarly to their peers.⁵ Here, the citizen obtains an additional benefit when matching their support action to compliance. Alternatively, the citizen may be concerned with backsliding, conceptualized as executive aggrandizement (see Grillo et al., 2024). Aggrandizement occurs with some probability when the leader is strong. In this case, the citizen incurs a cost when the consequence is realized. We show that, when comparing the citizen’s behavior conditional on believing in the *BO* narrative or the *SN*, the citizen’s core calculus remains unchanged relative to the baseline model. Under rational expectations, it is only the leader’s base that supports: the citizens who obtain sufficiently large ideological benefits. Compared to the baseline model, this base

⁵More generally, citizens want to coordinate with others, as in e.g., Tyson (2018). However, given that in our model, only the support decision is endogenous, not the compliance decision by non-strategic citizens, we do not use the term coordination.

is larger in the case of compliance concerns and smaller in the case of aggrandizement. When the citizen believes in the *SN*, the leader enjoys more support, similar to the baseline model. For intermediate levels of ideological benefits, there exist multiple equilibria that differ in the likelihood of support.

We then investigate which narrative is most desirable for the leader, focusing on narratives that change at most one link from the DAG that represents the true data generating process.⁶ Consider first the case when the citizen gets a benefit when matching their support action to compliance. Here, the *SN* is again the most desirable for the leader: it increases the subjective assessment that high economic performance can be obtained and also acknowledges that support increases leader strength, which leads to compliance by other citizens. The same result holds when the citizen incurs a *small* cost when democratic backsliding occurs. By contrast, when there are *large* costs to democratic backsliding, a narrative which portrays it as exogenous is most desirable for the leader. The intuition is that such a “blame-shifting” narrative breaks the perceived link between citizen support and backsliding. The *SN* concedes the existence of this link, rendering it less desirable for the leader. Empirical research has found that in a number of contexts, citizens’ concerns about backsliding are relatively small, however (e.g., Svulik, 2020; Graham and Svulik, 2020). This suggests that in empirically plausible scenarios, the *SN* is the narrative that is most desirable for political leaders.⁷

Our motivation for modeling the citizen’s perception in terms of causal processes is familiar to political scientists. Causality is difficult to establish empirically; it is often impossible with observational data. We consider beliefs about plausible data generating processes that guide the citizen’s decision-making. Moreover, we focus on a citizen who takes a narrative as given but is otherwise perfectly rational, making optimal choices conditional on beliefs about the consequences of her actions. Given that data generating processes are typically unknowable without experimentation, this assumption resonates

⁶The number of possible narratives as a function of nodes increases fast: without restrictions, 543 DAGs can be produced with only four nodes. Because there are no existing methods to find the set Personal Equilibria of each DAG easily, we calculate them individually. This precludes analyses of more complicated data generating processes with many nodes.

⁷But see Druckman (2024) and Schlipphak et al. (2023) on the importance of blame-shifting narratives in the context of democratic backsliding.

with the complexity of the real world. Our contribution is to show that even if the *SN* is not the true data generating process, it can still influence decisions due to misattribution of causes and effects. As a result, believing in the *SN* has major implications for the support (anti-democratic) politicians enjoy, the incentives to influence public opinion, and democratic resilience. More broadly, we emphasize the applicability of the current framework to the theoretical and empirical study of other political phenomena.

Our paper primarily contributes to the literature on how political leaders maintain public support, especially in semi-authoritarian contexts. Scholars have examined the role of cooptation and the provision of goods (Bueno De Mesquita et al., 2005; Gandhi and Przeworski, 2006), elections (Little et al., 2012; Luo and Rozenas, 2018), repression (Rozenas and Stukal, 2019; Tyson, 2018), censorship (Shadmehr and Bernhardt, 2015), and propaganda (Gehlbach and Sonin, 2014; Edmond, 2013; Chen and Xu, 2017; Little, 2017). Our key contribution is to analyze the effectiveness and correlates of the *SN*—a causal story that is significantly more complex than simply persuading the citizen that a fact is true or false (Öztürk, 2023). Narratives such as the *SN* are especially common in the context of democratic backsliding (for a recent review, see Grillo et al., 2024). A central puzzle in this literature is why voters support politicians who openly agitate against democratic institutions (Luo and Przeworski (2023); Acemoglu, Robinson and Torvik (2013); Chiopris, Nalepa and Vanberg (2025); see also Miller (2021); Howell et al. (2023); Horz (2021); Helmke, Kroeger and Paine (2022); Svulik (2020)). Previous work has proposed voters’ concerns about the economy as one key reason (Gratton and Lee, 2024). Consistent with the *SN*, anti-democratic politicians often argue that strong leadership—in the sense of few formal or informal restrictions on their rule—leads to better economic performance. Similar to Grillo and Prato (2023), we emphasize that expectations play a key role in explaining whether citizens support the leader.

We also complement the literature on belief formation. While classic rational choice work takes beliefs as given, recent research explores the causes and consequences of endogenous beliefs (e.g., Minozzi, 2013; Bénabou and Tirole, 2006; Ashworth, Fowler et al., 2020; Lockwood, 2017). Perhaps most closely related is Bénabou and Tirole (2006).

There, citizens choose a level of effort to increase their income. Returns to effort are influenced by an unknown state of the world. Citizens receive a signal about this state and choose their “recall rate,” i.e., the probability with which they can accurately recall a bad signal when choosing effort. Bénabou and Tirole (2006) show that biased beliefs can be optimal by functioning as a commitment device. Hence, similar to our paper, biased beliefs about how their economic fortunes are determined make citizens choose an action they would otherwise not choose. Moreover, multiple, self-fulfilling equilibria are present in each model. However, the sources of bias in the two papers are different: in Bénabou and Tirole (2006), beliefs about an exogenous state of the world are biased because of time-inconsistent preferences; in ours, beliefs about the consequences of a citizen’s action (absent a state of the world) are biased because of the challenges associated with drawing causal inferences. Bénabou and Tirole (2006) is a demand-side theory of biased beliefs: the misperceptions arise from citizens’ own needs. We instead focus on narratives that are desirable for a leader. Our paper builds on the framework proposed by Spiegel (2016), whose distinct focus is on causal relationships in a multivariate context.⁸ This framework has seen a number of recent applications in the context of policy choice in democracies, including Eliaz and Spiegel (2020) and Spiegel, Eliaz and Galperti (2024) (see also Eliaz and Spiegel (2024) and Levy, Razin and Young (2022)). However, while scholars have investigated the role of the elites’ beliefs on the economy and political regimes (Abramson and Montero, 2020; Albertus and Gay, 2017), our paper focuses on citizens’ beliefs in different narratives and these narratives’ effectiveness in eliciting support for leaders.

Baseline Model

Setup

We present a model with two main components: a representative citizen, C , who chooses whether to support a leader, and a “historical database” to which she has access. We

⁸Little (2019) and Izzo, Martin and Callander (2023) also consider multivariate environments but focus on heterogeneous preferences (Little, 2019) and ideological competition (Izzo, Martin and Callander, 2023) instead.

refer to this database as the polity’s history, as it contains past choices and outcomes. The citizen uses data from the polity’s history to form beliefs about the likelihood of future outcomes. An example of a historical database is depicted in Table 1. This allows the computation of the marginal distributions of all variables—including actions of past decision-makers—and there is no uncertainty about these distributions (i.e., there are infinitely many observations).

As noted, the citizen chooses whether to support the leader, $a = 1$, or not, $a = 0$. For example, $a = 1$ could correspond to voting in favor of a referendum that would expand the leader’s powers, voting for his political allies, attending pro-government rallies or participating in counter-protests against regime opponents. After the support decision, two variables are realized: leader strength, $\theta \in \{0, 1\}$, and economic performance, $y \in \{0, 1\}$. We say that the leader is strong if $\theta = 1$ and economic performance is high if $y = 1$. Withholding support is not equivalent to removal from office; we assume that the leader remains in office regardless of the citizen’s action.

We interpret leader strength as a combination of institutional and political factors that allow for greater policy discretion. In the baseline model, the citizen does not care intrinsically about these policies and a strong leader does not represent democratic backsliding—we examine such considerations in an extension. As such, the baseline model is also applicable to electoral politics in consolidated democracies. However, the citizen does care about economic performance, which may correspond to low unemployment, low inflation, or high growth. Her payoff is:

$$u_C = y + ax, \tag{1}$$

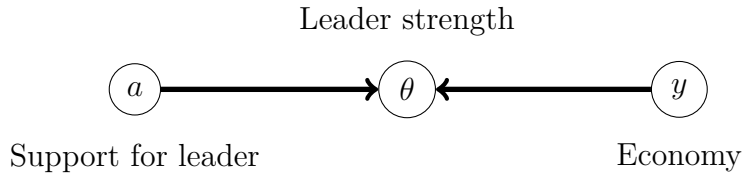
where $x \in \mathbb{R}$ is a parameter that governs the citizen’s relative ideological inclination to support the leader.⁹ The ideological benefits of supporting the leader, x , are exogenously fixed.

Consistent with theories emphasizing the ease with which leaders can buy off potential

⁹For example, suppose that the leader’s policy platform is x_I and there is a challenger whose position is x_C . Then, assuming linear loss, x is the relative ideological distance of the decisive citizen to the leader versus the challenger: $x = |x_I - x_V| - |x_C - x_V|$, where x_V is the voter’s ideal point.

challengers, and to show that the *SN* can elicit support even when it is incorrect, we assume that the true data generating process is the first DAG in Figure 1: citizen support and economic performance jointly affect leader strength, but economic performance is exogenous and high with probability $\gamma \in (0, 1)$.

The *BO* DAG



The *SN* DAG

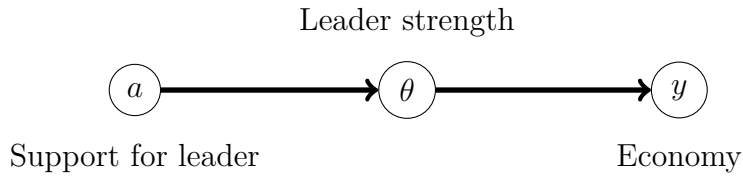


Figure 1: Two data generating processes linking citizen support (a), leader strength (θ), and economic performance (y): the *BO* DAG and the *SN* DAG.

Both DAGs in Figure 1 are entirely non-parametric. For simplicity, we assume that support and strong economic performance are jointly necessary and sufficient for leader strength: $\theta = ay$. In other words, the leader is strong if and only if both the economy is good *and* the citizen supports him. This specification implies that citizen support and economic performance are complements. Later, we consider a more general specification.

The *Strongman Narrative* is given in the lower part in Figure 1. It inverts the causal relationship between leader strength and the economy: leader strength causes good economic performance.

Solution Concept

We now describe our solution concept and how beliefs about causal processes influence decision-making. Following Spiegler (2016), our solution concept is Personal Equilibrium. In essence, in a Personal Equilibrium, citizens' beliefs about support must be consistent with the support those beliefs generate. We briefly review the general approach before

applying it to our environment.

Let $z = (z_i)_{i=1,\dots,n}$ be the collection of variables under consideration. Define $z_1 = a$ as the variable equal to the action of the decision-maker and A as the corresponding choice set. $(z_i)_{i=2,\dots,n} \equiv \tilde{z}$ are potentially payoff-relevant variables. The joint distribution of z is denoted by $\Pr(z)$. We use the notation \Pr for objective probabilities and, given any DAG R , \Pr_R for subjectively derived probabilities (e.g., \Pr_{BO} and \Pr_{SN} are subjective probabilities when employing the *BO* and *SN* DAGs, respectively). This subjective joint probability of a set of events $\Pr_R(z)$ is calculated by multiplying the probabilities of each event conditional on their causes, i.e., by using the following factorization formula:

$$\Pr_R(z) = \prod_{i=1}^n \Pr(z_i \mid z_{R(i)}), \quad (2)$$

where $R(i)$ is the set of direct parents of the node i (in the DAG R). Root nodes—events that are exogenous and not caused by other events in the DAG—are included unconditionally. All relevant beliefs can be deduced from the joint distribution $\Pr_R(z)$ using the usual probability operations, as we explain below. Because $\Pr(z_i)$ is generically not equal to $\Pr(z_i \mid z_j)$, different causal models lead to different inferences.

These concepts are familiar to social scientists. DAGs are often used to represent causal theories in an intuitive way (e.g., Morgan and Winship, 2015). They have also been used to make sense of learning as a realistic psychological model (Gopnik et al., 2004). An actor’s causal theory tells her *which variables to condition on* when forming expectations about quantities of interest. This operation is critical to empirical inquiry. For example, regression is essentially an operation to approximate the conditional expectation function (Angrist and Pischke, 2008). The innovation of Spiegel’s (2016) framework is to link these beliefs with expected utility maximization.

Once beliefs are formed, the citizen computes her subjective expected utility for each action and chooses the action that maximizes it. The expected utility may vary with the long-run frequency of choosing a specific action. This necessitates the following equilibrium approach:

Definition 1. (*Personal Equilibrium (Spiegler, 2016)*). Fix an arbitrary DAG R . A distribution $p \in \Delta(z)$ with full support on the choice set A is an ϵ -perturbed Personal Equilibrium if

$$a \in \arg \max_{a'} \sum_{\tilde{z}} Pr_R(\tilde{z} | a) u(a', \tilde{z})$$

whenever $p(a) > \epsilon$ and where, Pr_R is computed via the formula in (2). A distribution p^* is a Personal Equilibrium if there exists a sequence $p^k \rightarrow p^*$ of perturbations of p^* , as well as a sequence $\epsilon^k \rightarrow 0$, such that p^k is an ϵ^k -perturbed Personal Equilibrium for every k .

The citizen's subjective beliefs, computed via the factorization formula in Equation (2), should be well-defined (hence the ϵ -perturbation) and her equilibrium action must maximize her expected utility given those beliefs. Furthermore, the action probability must be a steady-state, i.e., the current action must be consistent with its long-run frequency. We discuss the practical steps on how to compute a Personal Equilibrium below.

Discussion

Before continuing with the analysis, we discuss several key features of the model. First, while for concreteness we refer to our variables as support (a), leader strength (θ), and economic performance (y), other interpretations are possible. For example, the citizen's action a can be volunteering for the leader or not protesting against a policy he champions. However, it is important to emphasize that the leader remains in office regardless of support. Hence, the support decision is akin to a referendum or a support choice that renders a policy choice more likely without affecting who is in power. Similarly, y may be thought of as another outcome that the citizen cares about, such as lower crime or better public health. The important condition is that this outcome must affect leader strength θ .

Second, we discuss how our approach relates to standard approaches. In most formal theories of politics, actors can correctly anticipate the consequences of their actions. In

particular, they know the data generating process. In models of political selection, for example, protesting or voting against the leader is a gamble that can pay off when the new officeholder is better than the leader (Ashworth, 2012). Thus, while citizens may face uncertainty (i.e., they may have to form beliefs about specific states of the world), they know the probability distribution over the consequences of their actions. In our account, citizens may have misperceptions about the effects of their actions, because they form their beliefs using the chain rule in Expression (2). We assume that citizens are otherwise rational—they have perfect recall, can calculate conditional probabilities and expected utilities conditional on believing in any data generating process, etc.—to isolate the effect of causal (mis-)perceptions.

Third, we focus on the effects of narratives on beliefs about the consequences of actions and behavior, consistent with empirical work that demonstrates the effectiveness of pro-regime messages across contexts (e.g., Yanagizawa-Drott, 2014; Carter and Carter, 2021; Adena et al., 2015). Our approach is complementary to formal work on communication that aims to understand the conditions under which a sender’s messages can shape behavior, assuming that receivers critically evaluate these messages (e.g., Gehlbach and Sonin, 2014). These models typically feature a relatively simple decision stage: the receiver takes the propagandist’s preferred action whenever the posterior belief regarding the relevant state of the world is sufficiently high. By contrast, while our decision stage is also relatively straightforward, the receiver’s decision-making is more complex, because past behavior influences the incentives for current actions.

Fourth, it is important to discuss our model’s notion of time. As noted, the historical database features infinitely many observations. When solving the model, the citizen computes a best response support action to the long-run frequency of support. The solution of the model describes a steady state in which the past and current actions are consistent with each other. Hence, there is no beginning of time and no end of time, but actions have to be consistent across periods. As a result, the entire history of support is endogenous. We return to this issue in the conclusion section.

Analysis

Finding a Personal Equilibrium means that we look for a number, $\beta^* \in [0, 1]$ that represents the *steady state probability of support* with which the citizen supports the leader.¹⁰

The citizen can observe past outcomes. Because there are infinitely many periods, the citizen knows the marginal distributions of all the variables in the model: γ is the frequency of high economic performance, and we denote by β the long-run frequency of support the leader enjoys. Finally, because $\theta = a\gamma$, the historical frequency of strong leadership is the product of these: $\gamma\beta$. An example history is shown in Table 1.

Table 1: The Historical Database containing three Variables

Index	Support a	Economic performance y	Leader strength θ
1.	1	1	1
2.	0	1	0
3.	1	0	0
\vdots	\vdots	\vdots	\vdots
Marg. Prob.	β	γ	$\beta\gamma$

To solve for a Personal Equilibrium, given beliefs about the data generating process, we proceed in three steps. The first step is to derive the *marginal* distributions of the variables that the citizen cares about. For the baseline case, the only non-trivial variable is economic performance, y —support a is a choice variable while leader strength θ does not affect the citizen’s utility. Equation (1) implies that given a narrative R and a long-run frequency of support β , the citizen chooses to support if:

$$EU_R(a = 1, \beta) \geq EU_R(a = 0, \beta) \iff \Pr_R(y = 1 \mid a = 1, \beta) + x \geq \Pr_R(y = 1 \mid a = 0, \beta). \quad (3)$$

To obtain the marginal distributions $\Pr_R(y = 1 \mid a, \beta)$ (that may depend on the long-run frequency of support, β), we first use the factorization formula in Equation (2)—given a narrative R —to obtain the joint distribution of all variables. From the joint distribution, we then sum over the other variables to obtain the marginal distribution of y .

The second step is to plug these probabilities into the expected utilities of each action

¹⁰In Definition 1, a Personal Equilibrium is a joint distribution over all variables in the model, but the distribution over the choice variable a is the key quantity (see Spiegler, 2016).

(as in Expression (3)). Because there are infinitely many historical observations in the database, the current action does not change the long-run frequency of support, nor associated beliefs. Hence, the citizen does not consider the effect of his action on β . The citizen's calculus defines the citizen's best response to the long-run frequency of support, β .

Finally, we impose an equilibrium requirement specifying that the expected utility calculation and the long-run frequency of support must be consistent with each other. This means that in an equilibrium with positive support, the net expected utility of support must be weakly positive.¹¹

Correct Expectations

We start with a benchmark analysis of the model, supposing that the citizen believes in the actual data generating process. She knows economic performance is exogenous and that it—alongside her decision whether to support—determines leader strength. Applying the factorization formula, the joint distribution of the variables a , θ , and y is:

$$\Pr_{BO}(a, \theta, y) = \Pr(a) \Pr(y) \Pr(\theta | a, y). \quad (4)$$

The joint distribution follows directly from the DAG: support (a) and economic performance (y) are exogenous, while leader strength (θ) depends on both of these factors.

The citizen cares about the marginal distribution of high economic performance, $y = 1$, conditional on her support decision, a , as outlined in Expression (3). To calculate $\Pr_{BO}(y = 1 | a, \beta)$, the citizen must sum over instances of leader strength:

$$\Pr_{BO}(y = 1 | a, \beta) = \sum_{\theta} \Pr(y = 1) \Pr(\theta | a, y).$$

Because $\Pr(y = 1) = \gamma$ and $\sum_{\theta} \Pr(\theta | a, y) = 1$, the expectation is just γ —economic

¹¹To understand the equilibrium concept, a comparison with Perfect Bayesian Equilibrium (PBE) might help. In a PBE, beliefs must be derived from Bayes' rule whenever possible and must be consistent with the sender's strategy. Here, beliefs are derived from the modified chain rule in Expression (2) and consistent with the citizen's strategy.

performance does not depend on support, a (nor its long-run frequency, β). Thus, from Expression (3), the citizen's expected payoff of supporting is $\gamma+x$ and the expected payoff of not supporting is γ . This means the citizen's choice only matters via the ideological payoff: it has no effect on the economy, and although it may effect leader strength, this is of no concern to the citizen. Here, the citizen's best response is equal to her equilibrium strategy: the citizen supports with steady state probability 1 if $x \geq 0$.

Incorrect Expectations: Believing in the SN

Best Response of the Citizen

In contrast to the case of rational expectations, the joint distribution over the variables a , θ , and y , as factored by the SN is

$$\Pr_{SN}(a, \theta, y) = \Pr(a) \Pr(\theta | a) \Pr(y | \theta).$$

By Expression (3), the citizen cares about the marginal distribution of y , conditional on the support choice a . This is equal to:

$$\Pr_{SN}(y = 1 | a, \beta) = \sum_{\theta} \Pr(\theta | a) \Pr(y = 1 | \theta). \quad (5)$$

Two conditional beliefs determine this expression: the conditional expectation of leader strength θ given action a and the conditional expectation of economic performance y given leader strength θ . The SN tells the citizen how to make sense of the polity's history: estimate the probability of leader strength conditional on the choice of support, $\Pr(\theta | a)$, and the probability of economic performance conditional on leader strength, $\Pr(y | \theta)$. Importantly, although a and y are independent, they are not independent *conditional* on θ . Thus, the citizen infers an indirect effect of her action on economic performance.

Note that $\Pr(y = 1 | \theta)$ is not well-defined if $\Pr(\theta)$ is 0. Specifically, when the citizen never supports, then the regime is always weak, and $\Pr(y = 1 | \theta = 1)$ is undefined. We

suppose that the database contains some observations of support. This ensures that the long-term frequency of citizen support is positive, $\beta > 0$ (see Table 1), and $\Pr(y = 1 \mid \theta = 1)$ is well-defined. Later, we make use of the perturbations in Definition 1 to consider the case in which the citizen never supports.

Given that the citizen believes in the *SN* and $\beta > 0$, the expected utility of supporting the leader, $\Pr_{SN}(y = 1 \mid a = 1) + x$, is:

$$\begin{aligned} & \sum_{\theta} \Pr(\theta \mid a = 1) \Pr(y = 1 \mid \theta) + x \\ &= \Pr(\theta = 1 \mid a = 1) \Pr(y = 1 \mid \theta = 1) + \Pr(\theta = 0 \mid a = 1) \Pr(y = 1 \mid \theta = 0) + x. \end{aligned}$$

The second line expands the sum in the first line. The next step is to compute these probabilities from the polity's history. Conditional on support, the leader is strong with probability γ and weak with probability $1 - \gamma$ (since $\theta = ay$ in the true data generating process). The *SN* asserts that support is the only cause of leader strength. This misleads the citizen to infer that γ —the exogenous probability of high economic performance—corresponds to the effect of her support on θ .

The citizen uses Bayes' rule to calculate the expectation of high economic performance when the leader is strong or weak:

$$\begin{aligned} \Pr(y = 1 \mid \theta = 1) &= \frac{\Pr(y = 1) \Pr(\theta = 1 \mid y = 1)}{\Pr(\theta = 1)}, \\ \Pr(y = 1 \mid \theta = 0) &= \frac{\Pr(y = 1) \Pr(\theta = 0 \mid y = 1)}{\Pr(\theta = 0)}. \end{aligned}$$

When the leader is strong—the first line—, this expression is equal to 1: the citizen always expects high economic performance conditional on a strong leader. The reason is that leader strength is in fact a *sufficient* condition for high economic performance. When the leader is weak, however, the citizen expects high economic performance with a lower probability: $\Pr(y = 1 \mid \theta = 0) = \frac{\gamma(1-\beta)}{1-\beta\gamma}$. Importantly, this expression is a decreasing function of the historical frequency of support, β . Higher frequency of past support leads to fewer incidences of leader weakness, and those periods are more likely to coincide

with a weak economy. Put differently, more historical support means fewer periods with a strong economy and a weak leader, which lowers the expectations of high economic performance when the leader is weak. Figure 2, left panel, illustrates this relationship.

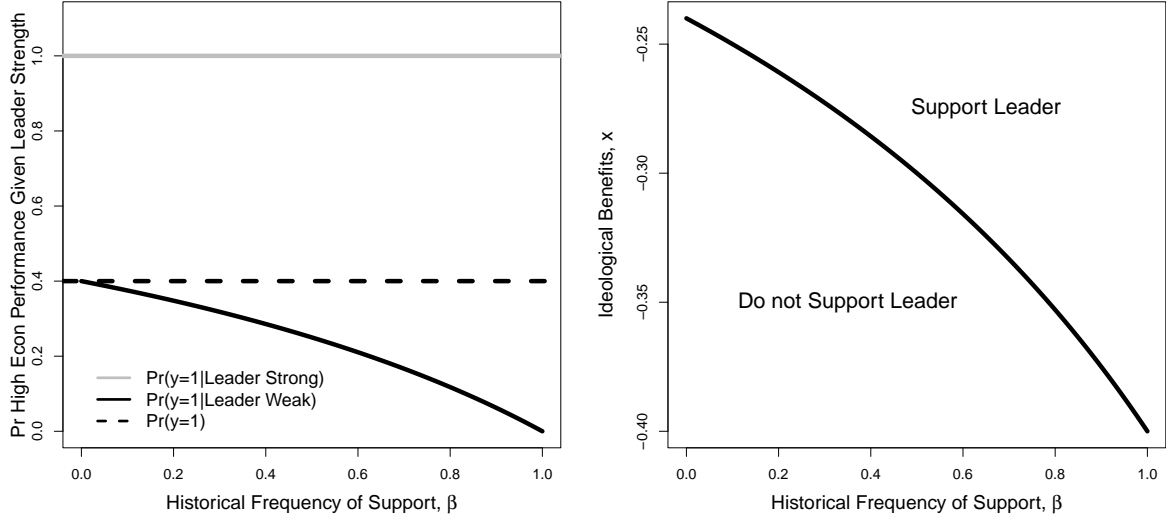


Figure 2: The left panel shows the key quantities $\Pr(y = 1 \mid \theta = 1)$ and $\Pr(y = 1 \mid \theta = 0)$ as a function of β in comparison to $\Pr(y = 1) = \gamma$, with $\gamma = 0.4$. The right panel shows the citizen's best response as a function of β and x .

Putting everything together, the expected utility of supporting the leader is:

$$\gamma 1 + (1 - \gamma) \frac{\gamma(1 - \beta)}{1 - \beta\gamma} + x.$$

Now consider the expected utility of not supporting. From Expression (5), when the citizen does not support, the leader is certainly weak. In this case, the citizen can expect high economic performance with probability $\frac{\gamma(1-\beta)}{1-\beta\gamma}$, as discussed above. Therefore, the expected utility of not supporting is $\frac{\gamma(1-\beta)}{1-\beta\gamma}$.

From these expected utilities, it follows that a citizen who believes in the *SN* supports the leader if and only if:

$$\underbrace{x + \frac{\gamma(1 - \gamma)}{1 - \beta\gamma}}_{NE(\beta, x)} \geq 0. \quad (6)$$

The left-hand side represents the net expected utility of supporting, which is a function of long-run support. This is equal to the net ideological benefit of supporting the leader

plus the perceived increase in the probability of obtaining high economic performance.

The first observation is that when the citizen believes that leader strength causes economic performance, she is willing to support a leader whose ideology she dislikes. This happens because moderates who are distant from the leader on ideological grounds nevertheless support because they (incorrectly) believe this helps obtain better economic outcomes. The second important observation from Expression (6) is that the citizen is more likely to support the leader, the higher the long-run frequency of support:

Lemma 1 (Historical complementarity). *The citizen's net expected utility of support is increasing in the long-run frequency of support, β .*

The intuition is that the correlation between leader strength and economic performance is stronger when there are more instances of past support. In the limit when citizens always support the leader, the state of the economy and the strength of the regime are perfectly correlated, and the current citizen believes that if she were to withhold support, this would ensure poor economic performance. At the other extreme when the citizen never supports the leader, leader strength and economic performance are uncorrelated in the data, because the former is always weak regardless of economic performance.

Notice that the citizen's error is one of *misattribution*. She believes that her support is the only cause of leader strength and estimates its effect size to be γ , which is in fact the exogenous probability of high economic performance. Moreover, she interprets the fact that economic performance is a *necessary condition* for a strong leader instead as the latter being a *sufficient condition* for the former. Although technically correct, this interpretation misattributes the causal link between the two variables. High economic performance is still possible under a weak leader, but the citizen estimates this probability to be $\frac{\gamma(1-\beta)}{1-\beta\gamma}$, which is strictly lower than the actual probability of γ whenever $\beta > 0$. Higher values of historical support β lead the citizen to believe that high economic performance is increasingly unlikely when the leader is weak—driven by the higher historical correlation between $\theta = 1$ and $y = 1$. Indeed, in the limit when $\beta = 1$, the citizen believes that leader strength is both necessary *and* sufficient for high economic performance. Given the *SN*, the above inferences are perfectly consistent with observed

data—any history produced by the *BO* DAG can be incorporated into the *SN* without raising doubts about the actual data generating process.

Personal Equilibrium

Recall that the expected utility function of the citizen contains β , the long-run frequency of support. This is because the citizen takes this quantity as given.¹² In equilibrium, the probability of current support must coincide with the long-run frequency of support. This requirement implies that for intermediate values of ideological benefits x , multiple equilibria exist:

Proposition 1. *There exist thresholds $x_0^{SN} \equiv -\gamma$ and $x_1^{SN} \equiv -\gamma(1-\gamma)$, such that there is positive support in equilibrium only if $x \geq x_0^{SN}$. When $x > x_1^{SN}$, in the unique equilibrium, the citizen always supports. When $x \in [x_0^{SN}, x_1^{SN}]$, there exists an equilibrium where the citizen always supports ($\beta^* = 1$) and one where she supports with interior probability $\beta^* = \beta^I \equiv \frac{x+\gamma(1-\gamma)}{x\gamma}$.*

Here, we assume that there are instances of support in the historical database, which allows the citizen to properly compute all relevant probabilities. The analysis goes through for a no-support equilibrium because, as required in Definition 1, one can consider a slight perturbation in support. To see this, suppose that we have an equilibrium in which the citizen *never* supports. So $\beta = 0$ and $\Pr(y = 1 \mid \theta = 1)$ is undefined. It must be optimal for the citizen to choose $a = 0$. To compute the expected utility of support, one considers a perturbation in which $\beta > 0$, which, as we show above, renders all probabilities well-defined. In fact, not supporting the leader is then optimal whenever $x < x_1^{SN}$. When $x < x_0^{SN}$, this is the unique equilibrium outcome.

Figures 3 and 4 illustrate the analysis. Consider Figure 3 first. In the left panel, ideological benefits are small ($x < -\gamma$). Here, the function NE is negative for all β . As a result, the best response to any β is to not support, i.e., $\Pr(a = 0) = 1$. The only steady state, where β and $\Pr(a = 1)$ are consistent with each other is $\beta^* = 0$. In the right

¹²One can relate this to the one-shot deviation principle familiar from dynamic games: the actions in other “periods” are held fixed. Hence, when the citizen decides whether to support in the “current period,” this quantity does not change.

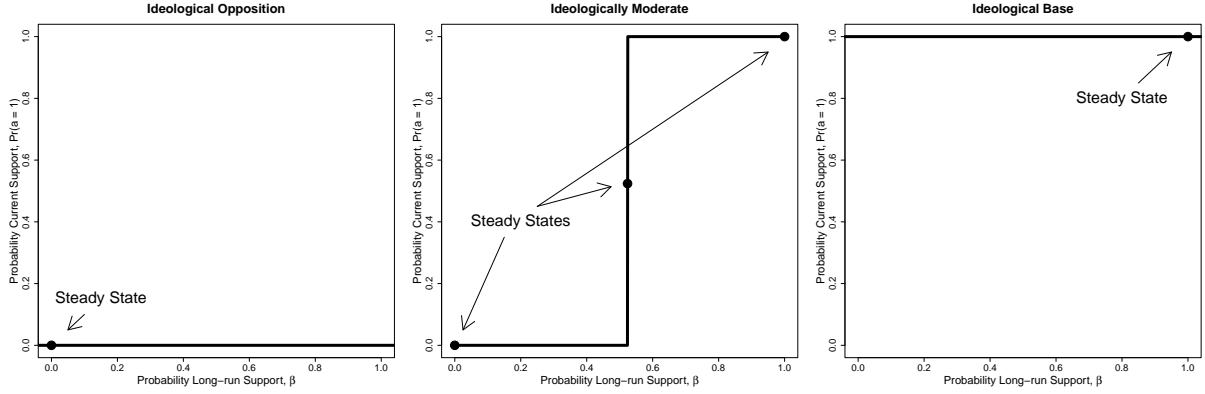


Figure 3: Best Responses and Steady States

panel, ideological benefits x are large ($x > -\gamma(1 - \gamma)$) and the function NE is positive for all β . As a result, the best response to any β is to support: $\Pr(a = 1) = 1$. The steady state in which β and $\Pr(a = 1)$ are consistent with each other is $\beta^* = 1$. Finally, the intermediate case ($x \in (-\gamma, -\gamma(1 - \gamma))$) is shown in the center panel in Figure 3. Here, for low levels of β , the best response is to not support, $\Pr(a = 1) = 0$. For high levels of β , the best response is to support, $\Pr(a = 1) = 1$. There is a critical level of β , $\frac{x + \gamma(1 - \gamma)}{x\gamma} \equiv \beta^I$, when the citizen is indifferent between supporting and not supporting. Thus, three steady states exist: $\beta^* \in \{0, \beta^I, 1\}$.¹³

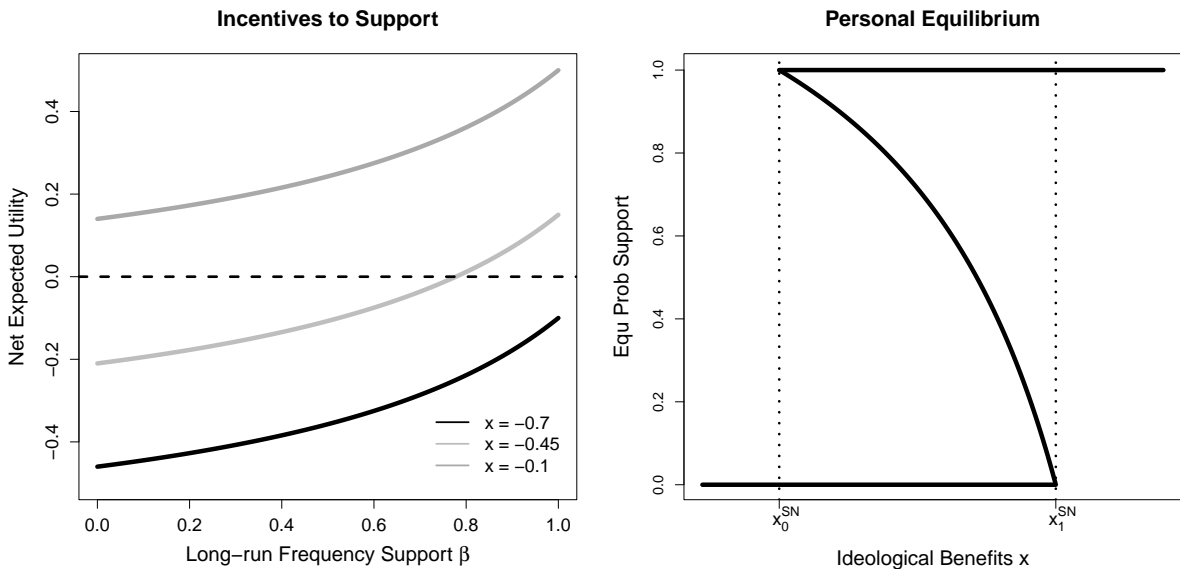


Figure 4: The citizen's net expected utility (left panel) and the personal equilibria of the model (right panel). Parameter values: $\gamma = 0.6$.

¹³Hence, in equilibrium, there are only three values in the historical data base: 0, β^I , or 1. Any other value of β is off-path.

Now consider Figure 4. For high levels of ideological payoffs, even if there was no support in the past, the ideological appeal is sufficiently high to render any perceived benefits of support on the economy irrelevant. The unique equilibrium then is to always support the leader. Similarly, for low levels of ideological payoffs, even concerns about the economy cannot overcome the citizen’s strong ideological dislike of the leader. Here, support cannot be sustained. Finally, in the intermediate region, the prevalence of past support is self-fulfilling: a higher frequency convinces the citizen that the positive effect of support on economic performance is large, rendering support today strictly preferred. When past frequency is low, the estimated effect of leader strength on economic performance—although positive—is low relative to the ideological distaste: support cannot be sustained. An interior frequency of past support sustains a mixed strategy equilibrium when it leaves the citizen exactly indifferent between the two actions. As usual with mixed strategy equilibria in games with strategic complements (e.g., Echenique and Edlin, 2004), this equilibrium is knife-edge and unstable—a small perturbation (in historical support) would cause the citizen to estimate a strictly positive or strictly negative net expected utility, resulting in a switch to one of the pure strategy equilibria. We ignore this equilibrium for the remainder of the paper.¹⁴

A substantive implication is that polities with similar conditions (in terms of ideological affinities to the leader and economic performances) can nevertheless be different in their willingness to support the leader—they form distinct “authoritarian legacies.” Besides differences in citizens’ utility functions, this effect contributes to our understanding of why similarly situated countries differ in the popular support their autocrats enjoy.

It is interesting to examine the effect of an increase in the exogenous probability of economic performance, γ , on equilibrium behavior. From Proposition 1, the following is immediate:

Corollary 1. *The necessary threshold for support, x_0^{SN} , decreases in γ while the sufficient threshold for support, x_1^{SN} , is minimized at $\gamma = 0.5$.*

¹⁴Note that this does not shrink the set of parameter values where the citizen supports the leader. The mixed-strategy equilibrium exists in the intermediate region where the ideological benefits are in the interval $[x_0^{SN}, x_1^{SN}]$. Below, we scrutinize the implications of selecting either pure-strategy equilibrium in this interior region.

Consider the threshold x_0^{SN} . Here, for low values of γ the expected returns to support are also low. This is because the citizen believes γ to correspond to the effect of her support on the leader strength. A low γ implies a weak inferred effect, which does little to encourage her. As γ increases, the citizen believes that her support is increasingly effective in ensuring leadership strength. The probability of high economic performance is very high when the leader is strong (recall that $\Pr(y = 1 \mid \theta = 1) = 1$). This helps the *SN* to elicit support from the citizen, making support easier to maintain.

Suppose next that the citizen supports if and only if $x > x_1^{SN}$. Here, γ has competing effects on the probability of support. On the one hand, a higher probability of high economic performance strengthens the perceived effect of support on leader strength, as above. For values of $\gamma < 0.5$, this effect dominates. On the other hand, a higher γ makes it more likely that the citizen also expects good economic performance when the leader is weak. This makes not supporting more attractive, which hurts the effectiveness of the *SN*. The latter effect is stronger when $\gamma > 0.5$.

Leader-Desirable Narratives

We have demonstrated above that given some ideological benefit x and a narrative R , the citizen either supports the leader or not. Naturally, for the leader, support is better—in this sense, a narrative R can be “successful.” We define this formally:

Definition 2. *A narrative R is successful for ideology x if it induces support (i.e., $\beta^* = 1$ in equilibrium) by a citizen with ideology x .*

This defines the success of a narrative as being able to elicit support for a given value of net ideological benefit. When the citizen has rational expectations, i.e., believes in the *BO* DAG $a \rightarrow \theta \leftarrow y$, the citizen’s steady state probability of support is 1 if and only if $x \geq 0$. Hence, the narrative *BO* is successful for any $x \geq 0$. By contrast, the *SN* is successful whenever $x \geq x^{SN}$, where $x^{SN} \in \{-\gamma, -\gamma(1 - \gamma)\}$ are the thresholds identified in Proposition 1.

It is clear that for some values of ideological benefits, x , the *SN* elicits support when

the *BO* cannot. As a result, the leader is better off under the *SN*. The next definition makes this precise:

Definition 3. *A narrative R is more desirable for the leader if it is successful for a larger set of ideologies in the sense of set inclusion.*

This definition says that a narrative that is successful for a broader set of ideological benefits is more desirable for the leader. As the above analysis shows, the *SN* is more desirable than the *BO*.

It is plausible that narratives that are more desirable for the leader are more likely to be chosen by the leader.¹⁵ However, note that we are engaging in a theoretical exercise here. The leader does not make an explicit choice, which is difficult to integrate into the current model setup, as there is no beginning and no end of time. We return to this issue in the conclusion section. For now, we wish to characterize which narrative is the most desirable for the leader. Are there other narratives that are better than the *SN* at eliciting support?

To do so, we first generalize a functional form assumption.¹⁶ The benchmark analysis above assumes that strong leaders can only arise when they have citizen support as well as high economic performance, i.e., $\Pr(\theta = 1) = ay$. We now allow strong leadership to emerge under a broader set of circumstances. Specifically, we generalize the leader strength technology to the following fully-interacted model:

$$\Pr(\theta = 1) = \alpha_0 + \alpha_a a + \alpha_y y + \alpha_i ay \equiv \alpha(a, y), \quad (7)$$

with $\alpha_i > 0$, $\alpha_0, \alpha_y, \alpha_a \geq 0$, and $\alpha_0 + \alpha_a + \alpha_y + \alpha_i \in [0, 1]$. Clearly, the baseline model is the special case where $\alpha_i = 1$. As a mild restriction, we assume that $\alpha_0 \alpha_i \geq \alpha_a \alpha_y$.¹⁷

¹⁵Besides choices by political elites, another explanation is that beliefs about the data generating process are transmitted via socialization or cultural institutions, as in Weber's traditional legitimacy (Weber, 2004).

¹⁶We continue to assume that the citizen fully believes the leader's narrative. In the Appendix, we consider the case in which there is *partial* believability of a narrative. All quantities must be adjusted by the probability with which the citizen believes a proposed narrative, but the fundamental forces remain in effect.

¹⁷Moreover, to guarantee that all posterior beliefs are well-defined for any long-run frequency of support β (i.e., without relying on perturbations), we sometimes assume that either α_0 or α_y is strictly positive.

With three nodes, there exist 10 possible non-degenerate DAGs in which a is a root. We show the following result:

Proposition 2. *Suppose leader strength is produced according to Equation (7). Then the SN is the most desirable narrative for the leader among all narratives with three nodes a , θ , and y .*

We now briefly discuss some of these other narratives. For example, there is another narrative in which a link from support to economic performance $a \rightarrow y$ is added. Here, it is effectively argued that lack of support (such as protesting) causes a decrease in economic performance. Everything else being equal, this is a “do not challenge” narrative common in authoritarian regimes. We show that this narrative does not work, leading to identical decision rules by the citizen as with rational expectations. By contrast, the SN does (weakly) increase support: we show in the Appendix that similar to Proposition 1, with the general leader strength technology in Expression (7), the SN elicits support whenever x is larger than one of the thresholds $\tilde{x}_0^{SN} < 0$ and $\tilde{x}_1^{SN} \leq 0$.

One might also suspect that the “do not challenge” narrative that argues for a direct link from a to y is especially potent when combined with the SN ’s promise that leader strength also delivers high economic performance. We call the resulting DAG the *working together* DAG, WT , displayed in Figure 5.

The WT DAG

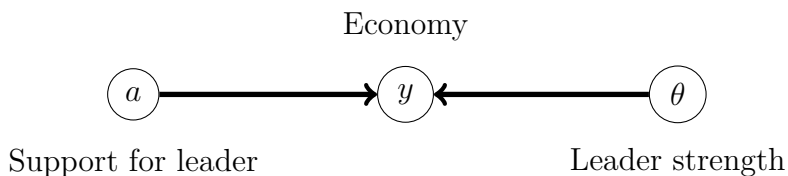


Figure 5: Another narrative linking citizen support (a), leader strength (θ), and economic performance (y): the WT DAG.

When the citizen believes in the WT narrative, the marginal probability of high economic performance conditional on some choice a is:

$$\Pr_{WT}(y = 1 | a) = \sum_{\theta} \Pr(\theta) \Pr(y = 1 | a, \theta).$$

This means that the narrative portrays leader strength as exogenous and argues that the citizen’s support action and strong leadership jointly produce high economic performance—they work in tandem.¹⁸ A citizen who believes the *WT* narrative supports if and only if:

$$\begin{aligned} & \Pr(\theta = 1) \Pr(y = 1 \mid a = 1, \theta = 1) + \Pr(\theta = 0) \Pr(y = 1 \mid a = 1, \theta = 0) + x \\ & \geq \Pr(\theta = 1) \Pr(y = 1 \mid a = 0, \theta = 1) + \Pr(\theta = 0) \Pr(y = 1 \mid a = 0, \theta = 0). \end{aligned}$$

Let $e_\theta = \Pr(y = 1 \mid a = 1, \theta) - \Pr(y = 1 \mid a = 0, \theta)$ denote the effect of support when regime strength is θ . Then, the citizen supports if:

$$\Pr(\theta = 1) (e_1 - e_0) + e_0 + x \geq 0.$$

The citizen here employs a difference-in-differences estimator, $e_1 - e_0$, and checks if this quantity, multiplied by the probability of a strong regime, is larger than $-x - e_0$, i.e., ideological costs minus the effect of support on economic performance when the leader is weak (e_0). Similar to the *SN*, we show in the Appendix that the *WT* narrative induces historical complementarity in the citizen’s best response and hence multiple personal equilibria. But in contrast to the *SN*, the *WT* does not expand the leader’s support base, it shrinks it—it is less desirable for the leader than either the *SN* or the *BO* narrative. The reason is that the *WT* narrative gives the citizen the wrong kind of agency—it asserts a direct link from her action to economic performance while omitting the *true* effect that support increases leader strength. Although support increases the probability of obtaining high economic performance when the leader is strong (due to the complementarity in the production function in Expression 7), it decreases performance when the leader is weak. The overall expected effect is negative.

Why is the *SN* so powerful? As we have seen, the *SN* is more successful in eliciting support than any other possible narrative involving the variables a , θ , and y . This is due

¹⁸For all probabilities $\Pr(y = 1 \mid a, \theta)$ to be well defined, we require $\alpha_0 + \alpha_y \gamma \in (0, 1)$ here, i.e., either α_0 or α_y must be strictly positive.

to its combination of three features. First, it (correctly) argues that the citizen’s action affects leader strength. This gives the citizen agency. Second, it reverts a causal arrow in the true data generating process and asserts that leader strength affects economic performance: the outcome the citizen cares about. Together, the first two points imply that the citizen can indirectly influence economic performance. Third, and more subtly, by not including a direct link between the citizen’s action and the outcome, the *SN* preempts a scrutinization of their correlation absent leader strength, thus preventing the citizen to realize that her action has, in fact, no effect on the outcome.

Other narratives may change the citizen’s behavior (because they affect the citizen’s beliefs about the consequences of her actions), but not to the leader’s advantage. For example, the *WT* leaves leader strength exogenous, thereby not exploiting the fact that the citizen’s support really does affect leader strength. Moreover, arguing that support has a direct effect on economic performance does not work. As argued above, this induces the citizen to analyze the historical database and find that the frequency of economic performance is the same, regardless of her support decision (i.e., $\Pr(y = 1 \mid a) = \gamma$ for any a).

Extension: A Richer Citizen Utility Function

So far, we have demonstrated that the *SN* is effective at eliciting citizen support and is indeed the most desirable narrative for the leader among all three-node narratives. We now check the robustness of our results when citizens face richer incentives. We focus on two concerns that the literature has identified as important: concerns to act similarly to other citizens (“compliance concerns”) and fears about aggrandizement.

Formally, suppose that the citizen cares about some consequence of leader strength, denoted by $c \in \{0, 1\}$. We assume that:

$$\Pr(c = 1 \mid \theta) = q\theta,$$

where q is some exogenous probability. One interpretation is that there is another group

of citizens who are non-strategic—they comply with probability q whenever the leader is strong. Another interpretation is that democratic backsliding happens with probability q when the leader is strong—for example, if there is an opportunity to engage in executive aggrandizement.¹⁹ In either case, the citizen’s utility function can be written as:

$$u_C = y + ax + c\lambda.$$

A positive $\lambda > 0$ can capture a desire to match the compliance action c by the non-strategic group of citizens.²⁰ If $\lambda < 0$, c can refer to backsliding, and the citizen suffers the disutility λ whenever this occurs (e.g., Grillo et al., 2024). Here, we distinguish between (institutional or more generally political) leader strength and executive aggrandizement, and we assume that leader strength causes aggrandizement.²¹ This distinction is motivated substantively: institutions that give more power to the current officeholder do not necessarily represent a degradation in democratic quality. However, an opportunity for aggrandizement, such as an international crisis, provides an opportunity to *exercise* the institution.

Finally, we continue to assume that the probability of a strong leader is given by the general specification in Equation (7). Our two main narratives are displayed in Figure 6.

Correct Expectations

We begin by analyzing a citizen who believes in the *BO* narrative. The expected payoff of supporting the leader is $\gamma + x + (\alpha_0 + \alpha_a + \gamma(\alpha_y + \alpha_i))q\lambda$, and the expected payoff of withholding support is $\gamma + (\alpha_0 + \alpha_y\gamma)q\lambda$. As before, the intuition is that the citizen knows that y is exogenous, and equal to 1 with probability γ . The citizen receives the

¹⁹To develop this a bit more, suppose that there is another variable $t \in \{0, 1\}$ that measures if there is an opportunity to engage in executive aggrandizement, and $\Pr(t = 1) = q$. We omit t from the data generating process to save on notation.

²⁰As mentioned above, this is motivated by the observation that in authoritarian contexts, citizens may want to coordinate with each other (e.g., Tyson, 2018).

²¹An alternative assumption is that aggrandizement causes strong leaders. In this specification, strong leaders can also come about when (exogenous) opportunities for aggrandizement are present. This is not implausible, but we focus on leader strength as a cause of backsliding here, i.e., leaders need to be strong to take advantage of an opportunity to initiate aggrandizement.

net ideological payoff x when she supports. Re-arranging, the citizen supports if

$$x \geq -(\alpha_a + \gamma\alpha_i)q\lambda. \quad (8)$$

The right-hand side is a higher (lower) threshold than the baseline model if $\lambda < 0$ ($\lambda > 0$) driven by concerns about executive aggrandizement (the desire to match support to compliance).

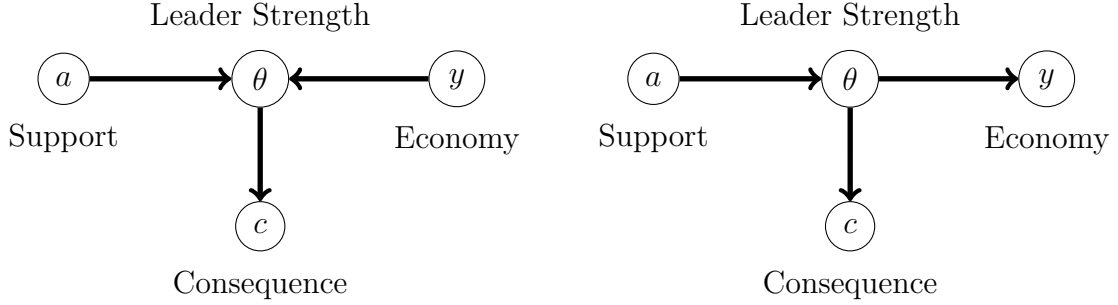


Figure 6: The *BO* and the *SN* DAG when there is an additional consequence (c) that the citizen cares about.

Incorrect Expectations: Believing in the *SN*

When the citizen believes in the *SN*, she needs to compute the marginal distributions of y and c . Consider y first. According to the *SN*, the marginal distribution of $y = 1$ is given by:

$$\Pr_{SN}(y = 1 | a) = \sum_{\theta, c} \Pr(\theta | a) \Pr(y = 1 | \theta) \Pr(c | \theta).$$

Suppose that the citizen supports the leader. Thus, we have $\Pr(\theta = 1 | a = 1) = \alpha_0 + \alpha_a + \gamma(\alpha_y + \alpha_i)$ and $\Pr(\theta = 0 | a = 1)$ with complementary probability $= 1 - \Pr(\theta = 1 | a = 1)$. The citizen needs to calculate the expectation of economic performance conditional on leader strength: $\Pr(y = 1 | \theta = 0)$ and $\Pr(y = 1 | \theta = 1)$. As before, we have:

$$\Pr(y = 1 | \theta = 1) = \frac{\Pr(y = 1) \Pr(\theta = 1 | y = 1)}{\Pr(\theta = 1)} \quad \text{and}$$

$$\Pr(y = 1 | \theta = 0) = \frac{\Pr(y = 1) \Pr(\theta = 0 | y = 1)}{\Pr(\theta = 0)}.$$

The marginal with respect to the consequence c is simply $(\alpha_0 + \alpha_a + \gamma(\alpha_y + \alpha_i))q$, as in the case where the citizen holds correct expectations.

Now consider the expected utility of not supporting. This is equal to:

$$\Pr(\theta = 1 | a = 0) \Pr(y = 1 | \theta = 1) + \Pr(\theta = 0 | a = 0) \Pr(y = 1 | \theta = 0) + (\alpha_0 + \alpha_y \gamma)q\lambda.$$

Re-arranging and simplifying, the citizen supports if:

$$\underbrace{[\Pr(\theta = 1 | a = 1) - \Pr(\theta = 1 | a = 0)]}_{\text{Effect of Support on Leader Strength}} \underbrace{[\Pr(y = 1 | \theta = 1) - \Pr(y = 1 | \theta = 0)]}_{\text{Effect of Leader Strength on Economy}} + \underbrace{(\alpha_a + \gamma\alpha_i)q}_{\text{Effect of Support on Consequence}} \lambda + x \geq 0.$$

This is the same as in the above analysis, except for the constant $(\alpha_a + \gamma\alpha_i)q\lambda$, the sign of which is determined by the sign of λ . Again, the game features historical complementarities, which can result in multiple Personal Equilibria, i.e., multiple steady state probabilities of support:

Proposition 3. *There exist thresholds $\tilde{x}_{\lambda,0}^{SN}$ and $\tilde{x}_{\lambda,1}^{SN}$, such that there is positive support in equilibrium only if $x \geq \tilde{x}_{\lambda,0}^{SN}$. When $x > \tilde{x}_{\lambda,1}^{SN}$, in the unique equilibrium, the citizen always supports. When $x \in [\tilde{x}_{\lambda,0}^{SN}, \tilde{x}_{\lambda,1}^{SN}]$, there exists an equilibrium where the citizen always supports ($\beta^* = 1$) and one where she supports with some interior probability.*

Leader-Desirable Narratives

Next, we again examine which narrative is the most desirable for the leader. We restrict attention to the set of narratives that are at most one link away from the DAG representing the true data-generating process. This includes many interesting cases. Consider the following two narratives, depicted in Figure 7. In both cases, the consequence c is exogenous; in the left panel, it is a cause of leader strength, θ ; in the right one, it is neither a cause nor a consequence. When $\lambda > 0$, so that the citizen has incentives to match his action a to c , the left panel represents a “bandwagon narrative” in which

the citizen is invited to “join” other citizens that support the leader. When $\lambda < 0$, so that the citizen is concerned about backsliding, both panels represent “blame-shifting” narratives in which the leader claims that leader strength does not cause backsliding. Rather, it happens exogenously due to other (unmodeled) forces (and it either causes leader strength or not).

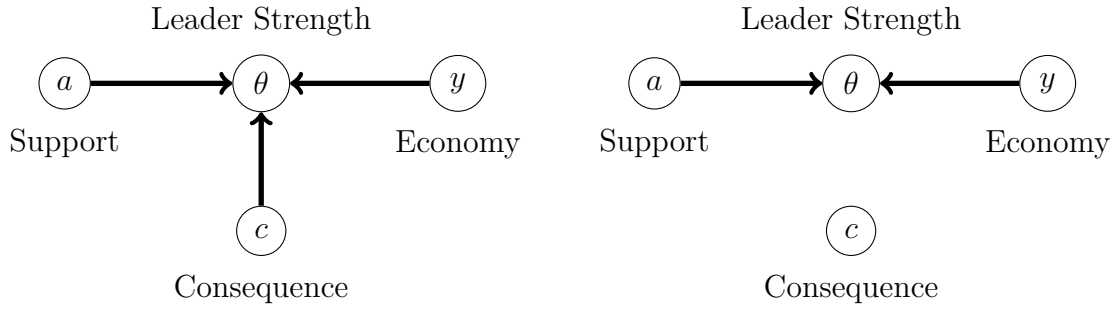


Figure 7: Narratives in which the consequence c is exogenous.

We can show the following result:

Proposition 4. *Consider the class of narratives that are at most one link away from the true data generating process. There are critical values $\bar{\lambda}_0$ and $\bar{\lambda}_1$, with $\bar{\lambda}_0 < \bar{\lambda}_1 < 0$, such that:*

- *when $\lambda > \bar{\lambda}_1$, the SN is the most desirable narrative for the leader.*
- *when $\lambda \in (\bar{\lambda}_0, \bar{\lambda}_1)$ and $\tilde{x}_{0,\lambda}^{SN}$ is the selected equilibrium threshold, the SN is the most desirable for the leader.*
- *when $\lambda \in (\bar{\lambda}_0, \bar{\lambda}_1)$ and $\tilde{x}_{1,\lambda}^{SN}$ is the selected equilibrium threshold, a narrative that portrays c as exogenous is most desirable for the leader.*
- *when $\lambda < \bar{\lambda}_0$, a narrative that portrays c as exogenous is most desirable for the leader.*

These cases are illustrated in Figure 8. Substantively, the result implies that unless the citizen has *severe* concerns about backsliding (relative to the economy and ideology), the SN is the most desirable narrative. Empirical analysis suggests that such concerns are unlikely in many contexts (e.g., Graham and Svovik, 2020; Svovik, 2020).

The intuition for Proposition 4 is as follows. The citizen is concerned with two future outcomes: economic performance, y , and the consequence, c . If the citizen is not strongly

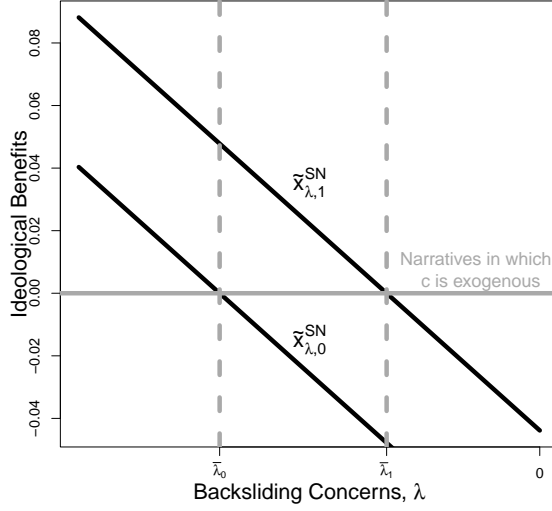


Figure 8: Thresholds for supporting for the SN and for narratives in which c is exogenous. The solid lines indicate equilibrium decision rules. The dashed lines indicate critical values, as identified in Proposition 4. Parameter values: $\alpha_0 = \alpha_y = \alpha_a = 0.1$, $\alpha_i = 0.3$, $\gamma = 0.4$, $q = 0.2$.

concerned with backsliding, the SN is again the narrative most desirable for the leader. The reason is that the SN is the only narrative (in this class) in which support increases the perceived probability of high economy performance. The SN is able to do this for the same reason as explained in the discussion of Proposition 2: it acknowledges the link between support and leader strength—thus giving the citizen agency—, asserts a causal link from leader strength to economic performance, and it does not include a *direct* link from support to economic performance. Rather, it induces the citizen to look at the conditional probability of economic performance when the leader is strong or weak. Hence, even when the SN is not the true data generating process, misattribution causes the citizen to assess a positive effect of support on economic performance.

By contrast, if the citizen has high costs for backsliding, the overwhelming desire is to avoid such an outcome. As a result, a narrative that breaks the link between the citizen's action, a , and backsliding, c , that—in the true data generating process—goes through leader strength, θ is more desirable for the leader. When c is exogenous (and either does not affect other variables or causes leader strength), the citizen neglects this link. She looks at the historical database and observes that backsliding happens with probability

$q \Pr(\theta = 1) = q(\alpha_0 + \alpha_a\beta + \alpha_y\gamma + \alpha_i\beta\gamma)$. While this depends on the long-run frequency of support, β , it does not change with the citizen's current support action, a . Hence, it does not affect the steady state probability of support in a Personal Equilibrium, and the citizen supports if ideological benefits are positive: $x \geq 0$. By contrast, the *SN* does acknowledge the link from leader strength to backsliding, rendering it less desirable under these conditions.²²

In sum, when the citizen cares both about economic performance and democratic backsliding, there are several narratives that are more desirable for the leader, compared to the case when the citizen has rational expectations, i.e., believes in the *BO* narrative. While there is no explicit choice for the leader, implicitly, there is a trade-off between the *SN* and blame-shifting narratives in which backsliding c is exogenous. It is worth noting, however, that in the situation when λ is negative, the leader would be better off if the narrative was the *SN* combined with a portrayal of backsliding as exogenous. Such a narrative is two links away and hence outside the scope of Proposition 4, although we do analyze it in the Appendix. We conjecture that for situations in which λ is negative, this narrative is the most desirable narrative for the leader within the class of narratives that are at most two links away from the *BO* DAG.²³

Conclusion

In their campaigns to vie for citizens' support, autocrats and would-be autocrats around the world emphasize the benefits of strong leadership and states via a *Strongman Narrative*. Why is the *SN* so prevalent around the world? Under which conditions does it successfully shape citizens' behavior? Analyzing a formal model of belief formation and support behavior, we emphasize that the effectiveness of the *SN* in eliciting support depends on history and ideological considerations. In particular, historical support for a strongman increases the incentives to support today. This is a form of historical com-

²²Consistent with this result, Druckman (2024) and Schlipphak et al. (2023) emphasize the importance of blame-shifting narratives in the context of democratic backsliding.

²³As noted below, an exploration of *all* narratives in complex environments, i.e., with many variables, is an important avenue for future research.

plementarity that results in multiple (Personal) Equilibria for intermediate levels of net ideological payoffs (i.e., for moderates). In many theories of autocratic politics, players' incentives are assumed to be such that their actions are strategic complements. For example, the incentive to protest is increasing in the number of players that are also protesting because it is then more likely that the regime is overthrown. In our model, the incentive to support is also increasing in past support, but this follows from narrative-specific belief formation, rather than assumed features of the environment. Finally, we show that the *SN* is the most desirable narrative for the leader among a large set of narratives—unless the citizen has sufficiently large costs of democratic backsliding. Then, a blame-shifting narrative is the most desirable for the leader.

In our analysis, we make a number of restrictive assumptions. In particular, we analyze a deterministic model in which the entire history of support is endogenous. Future work might relax this. Consider the following. Suppose in period 1 the citizen has access to some exogenous history. Denote the frequency of support in that history by β_0 . Given high enough x , this citizen will support the leader for some β_0 . Denote this endogenous, equilibrium probability of support in period 1 by β_1 . In period 2, the observed history consists of β_0 (exogenous) and β_1 (endogenous). Again, given high enough x , the citizen will support for some β_0 and β_1 . In fact, if x does not change across periods, given historical complementarity, whenever the citizen supports in some period t , he will support in all future periods. This resembles our analysis conducted here (except that it is easier to introduce actions by other players, such as a leader, given that there is a clear timeline with a start date). However, if the ideology changes across periods, equilibrium play is more interesting, because incentives to support are not monotonic in time. In the Appendix, we analyze a version of our model in which x (and therefore support) is probabilistic, while maintaining the assumption that the entire history of support is endogenous. We show that an equilibrium exists in this case as well, but is not necessarily unique. Specifically, equilibrium multiplicity occurs if there is sufficient probability mass in the interval in which there are multiple equilibria for a fixed level of ideological benefits. We also show that the probability of high economic performance, γ ,

ambiguously affects the equilibrium threshold—even if there is a unique equilibrium.

In this paper, we also restrict attention to pure-strategy personal equilibria. When we incorporate the mixed-strategy equilibrium into the analysis, Corollary 1 features more complicated comparative statics, because the citizen’s mixed strategy depends on γ . That said, the fact that the *SN* is the most desirable narrative for the leader for three variable DAGs does not depend on this selection. Finally, concerning the leader-desirability of narratives, we restrict attention to all three node DAGs and, with four nodes, to DAGs that are at most link away from the true DAG. Future work should build on these results to analyze a broader pool of possible narratives, and their relationship with other instruments, such as repression or cooptation, to maintain public support.

References

- Abramson, Scott F and Sergio Montero. 2020. “Learning about Growth and Democracy.” *American Political Science Review* 114(4):1195–1212.
- Acemoglu, Daron, Cevat Giray Aksoy, Ceren Baysan, Carlos Molina and Gamze Zeki. 2024. Misperceptions and Demand for Democracy under Authoritarianism. Technical report National Bureau of Economic Research.
- Acemoglu, Daron, James A Robinson and Ragnar Torvik. 2013. “Why do voters dismantle checks and balances?” *Review of Economic Studies* 80(3):845–875.
- Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa and Ekaterina Zhuravskaya. 2015. “Radio and the Rise of the Nazis in Prewar Germany.” *The Quarterly Journal of Economics* 130(4):1885–1939.
- Albertus, Michael and Victor Gay. 2017. “Unlikely Democrats: Economic Elite Uncertainty under Dictatorship and Support for Democratization.” *American Journal of Political Science* 61(3):624–641.
- Angrist, Joshua D. and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics*. Princeton, NJ: Princeton university press.
- Ashworth, Scott. 2012. “Electoral Accountability: Recent Theoretical and Empirical Work.” *Annual Review of Political Science* 15:183–201.
- Ashworth, Scott, Anthony Fowler et al. 2020. “Electoral versus Voters.” *Journal of Political Institutions and Political Economy* 1(3):477–505.
- Beissinger, Mark and Stephen Kotkin. 2014. *Historical legacies of communism in Russia and Eastern Europe*. Cambridge University Press.
- Bénabou, Roland and Jean Tirole. 2006. “Belief in a Just World and Redistributive Politics.” *The Quarterly Journal of Economics* 121(2):699–746.

- Bueno De Mesquita, Bruce, Alastair Smith, Randolph M. Siverson and James D. Morrow. 2005. *The Logic of Political Survival*. Cambridge, MA: MIT press.
- Carter, Erin Baggott and Brett L Carter. 2021. “Propaganda and protest in autocracies.” *Journal of Conflict Resolution* 65(5):919–949.
- Chen, Jidong and Yiqing Xu. 2017. “Information manipulation and reform in authoritarian regimes.” *Political Science Research and Methods* 5(1):163–178.
- Chiopris, Caterina, Monika Nalepa and Georg Vanberg. 2025. “A wolf in sheep’s clothing: Citizen uncertainty and democratic backsliding.” *The Journal of Politics* 87(4):1272–1287.
- Druckman, James N. 2024. “How to study democratic backsliding.” *Political Psychology* 45:3–42.
- Echenique, Federico and Aaron Edlin. 2004. “Mixed equilibria are unstable in games of strategic complements.” *Journal of Economic Theory* 118:61–79.
- Edmond, Chris. 2013. “Information manipulation, coordination, and regime change.” *Review of Economic Studies* 80(4):1422–1458.
- Elçi, Ezgi. 2022. “Politics of nostalgia and populism: Evidence from Turkey.” *British Journal of Political Science* 52(2):697–714.
- Eliaz, Kfir and Ran Spiegler. 2020. “A Model of Competing Narratives.” *American Economic Review* 110(12):3786–3816.
- Eliaz, Kfir and Ran Spiegler. 2024. “News Media as Suppliers of Narratives (and Information).” *arXiv preprint arXiv:2403.09155* .
- Frantz, Erica. 2018. *Authoritarianism: What Everyone Needs to Know®*. Oxford University Press.
- Gandhi, Jennifer and Adam Przeworski. 2006. “Cooperation, cooptation, and rebellion under dictatorships.” *Economics & politics* 18(1):1–26.

- Gehlbach, Scott and Konstantin Sonin. 2014. "Government Control of the Media." *Journal of public Economics* 118:163–171.
- Gopnik, Alison, Clark Glymour, David M. Sobel, Laura E. Schulz, Tamar Kushnir and David Danks. 2004. "A Theory of Causal Learning in Children: Causal Maps and Bayes Nets." *Psychological Review* 111(1):3.
- Graham, Matthew H and Milan W Svobik. 2020. "Democracy in America? Partisanship, polarization, and the robustness of support for democracy in the United States." *American Political Science Review* 114(2):392–409.
- Gratton, Gabriele and Barton E Lee. 2024. "Liberty, security, and accountability: The rise and fall of illiberal democracies." *Review of Economic Studies* 91(1):340–371.
- Grillo, Edoardo and Carlo Prato. 2023. "Reference points and democratic backsliding." *American Journal of Political Science* 67(1):71–88.
- Grillo, Edoardo, Zhaotian Luo, Monika Nalepa and Carlo Prato. 2024. "Theories of Democratic Backsliding." *Annual Review of Political Science* 27.
- Gurr, Ted. 1968. "A Causal Model of Civil Strife: A Comparative Analysis Using New Indices." *The American Political Science Review* 62(4):1104–1124.
- Gurr, Ted Robert. 1970. "Sources of Rebellion in Western Societies: Some Quantitative Evidence." *The ANNALS of the American Academy of Political and Social Science* 391(1):128–144.
- Helmke, Gretchen, Mary Kroeger and Jack Paine. 2022. "Democracy by deterrence: Norms, constitutions, and electoral tilting." *American Journal of Political Science* 66(2):434–450.
- Horz, Carlo M. 2021. "Electoral Manipulation in Polarized Societies." *The Journal of Politics* 83(2):483–497.

- Howell, William G, Kenneth A Shepsle, Stephane Wolton et al. 2023. "Executive Absolutism: The Dynamics of Authority Acquisition in a System of Separated Powers." *Quarterly Journal of Political Science* 18(2):243–275.
- Izzo, Federica, Gregory J Martin and Steven Callander. 2023. "Ideological Competition." *American Journal of Political Science* 67(3):687–700.
- Levitsky, Steven and Daniel Ziblatt. 2018. *How Democracies Die*. New York, NY: Broadway Books.
- Levy, Gilat, Ronny Razin and Alwyn Young. 2022. "Misspecified politics and the recurrence of populism." *American Economic Review* 112(3):928–962.
- Little, Andrew T. 2017. "Propaganda and credulity." *Games and Economic Behavior* 102:224–232.
- Little, Andrew T. 2019. "The Distortion of Related Beliefs." *American Journal of Political Science* 63(3):675–689.
- Little, Andrew T et al. 2012. "Elections, fraud, and election monitoring in the shadow of revolution." *Quarterly Journal of Political Science* 7(3):249–283.
- Lockwood, Ben. 2017. "Confirmation bias and electoral accountability." *Quarterly Journal of Political Science* 11(4):471–501.
- Luo, Zhaotian and Adam Przeworski. 2023. "Democracy and its Vulnerabilities: Dynamics of Democratic Backsliding." *Quarterly Journal of Political Science* 18(1):105–130.
- Luo, Zhaotian and Arturas Rozenas. 2018. "Strategies of election rigging: trade-offs, determinants, and consequences." *Quarterly Journal of Political Science* 13(1):1–28.
- Miller, Michael K. 2021. "A republic, if you can keep it: Breakdown and erosion in modern democracies." *The Journal of Politics* 83(1):198–213.
- Minozzi, William. 2013. "Endogenous Beliefs in Models of Politics." *American Journal of Political Science* 57(3):566–581.

- Morgan, Stephen L. and Christopher Winship. 2015. *Counterfactuals and Causal Inference*. Cambridge: Cambridge University Press.
- Öztürk, Aykut. 2023. “Whisper sweet nothings to me Erdoğan: developmentalist propaganda, partisan emotions, and economic evaluations in Turkey.” *Democratization* 30(7):1357–1379.
- Pop-Eleches, Grigore and Joshua A Tucker. 2017. *Communism’s shadow: Historical legacies and contemporary political attitudes*. Princeton University Press.
- Rozenas, Arturas and Denis Stukal. 2019. “How Autocrats Manipulate Economic News: Evidence from Russia’s State-Controlled Television.” *The Journal of Politics* 81(3):982–996.
- Schlipphak, Bernd, Paul Meiners, Oliver Treib and Constantin Schäfer. 2023. “When are governmental blaming strategies effective? How blame, source and trust effects shape citizens’ acceptance of EU sanctions against democratic backsliding.” *Journal of European Public Policy* 30(9):1715–1737.
- Shadmehr, Mehdi and Dan Bernhardt. 2015. “State censorship.” *American Economic Journal: Microeconomics* 7(2):280–307.
- Spiegler, Ran. 2016. “Bayesian Networks and Boundedly Rational Expectations.” *The Quarterly Journal of Economics* 131(3):1243–1290.
- Spiegler, Ran, Kfir Eliaz and Simone Galperti. 2024. “False Narratives and Political Mobilization.” *Journal of the European Economic Association* .
- Svolik, Milan W. 2020. “When Polarization Trumps Civic Virtue: Partisan Conflict and the Subversion of Democracy by Incumbents.” *Quarterly Journal of Political Science* 15(1):3–31.
- Tyson, Scott A. 2018. “The agency problem underlying repression.” *The Journal of Politics* 80(4):1297–1310.

Weber, Max. 2004. *The vocation lectures*. Indianapolis, IN: Hackett Publishing.

Yanagizawa-Drott, David. 2014. "Propaganda and Conflict: Evidence from the Rwandan Genocide." *The Quarterly Journal of Economics* 129(4):1947–1994.