"I Alone Can Fix It:" the Strongman Narrative and Democratic Backsliding*

Carlo M. Horz† Korhan Kocak‡

Abstract

Aspiring autocrats often vie for citizens' support to dismantle democratic institutions. One widely used instrument is the Strongman Narrative (*SN*): propaganda claiming that a strong leader can improve economic performance. We build a game-theoretic model to study the effectiveness and correlates of the *SN*. We find that the *SN* increases moderates' support for the leader, but the extent to which it does depends on the polity's history. In particular, we find a form of "authoritarian legacy:" past support for authoritarian leaders increases citizens' tendencies to support the leader today. In countries with an authoritarian legacy, the *SN* is most effective in times of prosperity, whereas without a legacy, it is more effective in times of volatility. We also study how the *SN* interacts with electoral manipulation. Under some conditions, manipulation and propaganda act as substitutes: leaders' incentive to manipulate elections are lower when the citizen believes in the *SN*.

*A previous version was circulated under the title "How To Keep Citizens Disengaged: Propaganda and Causal Misperceptions." We thank Quintin Beazer, Afiq bin Oslan, Federica Izzo, Holger Kern, Andrew Little, 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, Texas A&M University, College Station, TX 77843-4348. E-mail: carlo.horz@tamu.edu. Web: www.carlohorz.com.

‡Assistant Professor, Division of Social Science, NYU Abu Dhabi, PO Box 129188, Saadiyat Island, Abu Dhabi. E-mail: kkocak@nyu.edu. Web: www.korhankocak.com.

From Vladimir Putin to Recep Tayyip Erdogan and 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 after the end of the Cold War. Indeed, it has resulted in a number of instances of democratic backsliding or even collapse (Frantz, 2018; Levitsky and Ziblatt, 2018). One important tool these leaders used to expand their powers is a propagandistic narrative which likens their rule to a past when the country was strong—both politically and economically—under ostensibly benevolent yet decisive rulers, such as the Czars, Sultans, and Pharaohs (Elçi, 2022). Implicit in this narrative is a causal theory we term the *strongman narrative* (*SN*): a strong leader makes a strong economy more likely. Indeed, Putin, Erdogan, and Orban have won elections and experienced broad public support during their rule despite resorting to electoral manipulation and aggrandizement. Under which conditions is the *SN* effective at strengthening citizens' support for the leader? And how does it interact with incentives to manipulate elections?

In this paper, we answer these questions, modeling the *SN* as a causal graph, following Spiegler's (2016) innovation to analyze misperceptions using directed acyclical graphs (DAGs). Our baseline model introduces the core misperception. Our main model analyzes a more complex scenario in which backsliding can occur: elections may be manipulated and executive aggrandizement may take place. In both cases, we find that the *SN* increases the citizens' support for the leader, as it can sway ideological moderates. How much the *SN* can help the leader elicit support depends on the context. In particular, we demonstrate a form of "authoritarian legacy:"[1] the citizen today is more likely to support the leader when historical leaders enjoyed support in the past. Finally, we find that the *SN* has competing effects on optimal electoral manipulation: under some conditions, manipulation and propaganda are substitutes.

Our baseline model features a series of representative, identical citizens in an infinite horizon setting. In every period, a citizen decides whether to support the incumbent or not, observing past choices and outcomes—what we refer to as the polity's history. The

[1]Here, 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)

citizen has two concerns. First, she cares about ideology and wants to support a politician who is ideologically aligned. Second, she wishes to obtain high economic performance. We assume that economic performance is exogenous in the true data generating process, but 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. This is why we call the true data generating process the buy-off DAG (BO).

The citizen, however, may believe the true data generating process to be the Strongman Narrative. The SN asserts that citizens' support causes a strong leader, which in turn causes high economic performance. Hence, the SN is a causal guide to make sense of the polity's history, influencing the citizen's beliefs about the economic consequences of supporting the leader. It prescribes that the citizen must estimate the probability of high economic performance by conditioning on a collider: leader strength. Since the citizen knows that their support increases leader strength, this encourages her to support the leader, even if he is not attractive on ideological grounds.

We compare the citizen's behavior when she believes in each of the causal processes. If the citizen has the correct data generating process—the BO DAG—she only supports a leader whose net ideological benefit is positive. In other words, the leader's support consists of his ideological base. The SN can allow 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, in reality, a good economy is a necessary condition for leader strength. 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 true 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.

The above logic implies a *historical complementarity*, whereby 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, which is "sanctified through the unimaginably ancient recognition and habitual orientation to conform" (Weber, 2004, p. 34). Formally, it contrasts with existing formal theories of (semi-)authoritarian politics in which payoff externalities induce cross-sectional, strategic complements among similarly situated actors (Little, 2012; Fearon, 2011). While historical complementarity is different, it also breeds equilibrium multiplicity. In particular, we solve for a "personal equilibrium" in which the probability that the citizen supports today and its historical frequency are consistent with each other. We demonstrate that for a range of ideological benefits, the model features multiple equilibria that differ in the amount of support. This finding explains why aspiring autocrats in countries that look similar in terms of their material circumstances can enjoy different degrees of support.

We also study the effectiveness of the *SN*, defined as the difference in the support the leader enjoys when the citizen believes in the *SN* relative to the *BO*. This quantity is always positive: the leader enjoys more support under the *SN*. Moreover, the effectiveness of the *SN* depends on the (exogenous) probability of high economic performance. There are two competing effects. On the one hand, a higher probability of good economic performance strengthens the observed positive effect of support on leader strength. Because good economic performance and leader strength are positively correlated, a higher estimated effect of support on leader strength increases the effectiveness of the *SN*. On the other hand, a higher probability of good economic performance increases the chances of good economic performance even when the leader is weak, making support redundant. This effect decreases the effectiveness of the *SN*. We find that, in countries with an authoritarian legacy, the former effect dominates and citizens are most susceptible to the *SN* when the economy is performing well. In countries without an authoritarian legacy, in contrast, the *SN* is most effective in times of economic volatility.[2]

[2]As we detail below, we define countries with an authoritarian legacy as those in which the citizen supports when there are multiple equilibria.

We next extend the setup to include a richer set of outcomes that more explicitly focus on democratic backsliding. Specifically, we introduce two notions of backsliding: electoral manipulation and executive aggrandizement (see Grillo et al., 2024). We assume that the citizen is aware of electoral manipulation, which weakens the link between the citizen's choice of whether to support and who eventually holds office. The citizen is concerned about aggrandizement, which occurs with some probability when the leader has expanded powers. We show that the citizen's core calculus remains unchanged. 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 is smaller because the citizen is concerned with aggrandizement. When the citizen believes in the *SN*, the leader enjoys more support, similar to the baseline model. For intermediate levels of ideological benefits, multiple equilibria exist that differ in the likelihood of support. We also show that the effectiveness of the *SN* is ambiguous in the frequency with which the polity experiences good economic performance and, importantly, increasing in the level of electoral manipulation.

This last observation has important implications for our final model variation where we scrutinize how the citizen's beliefs interact with incentives to manipulate elections. Here, the leader chooses the level of manipulation, facing uncertainty about the voter's ideology. A higher level of electoral manipulation means that the probability of remaining in office is higher, even absent the citizen's support. Hence, electoral manipulation acts as an insurance policy, should the citizen choose to withhold support. As a result, precisely because the *SN* increases the leader's support, it makes electoral manipulation less necessary. When comparing optimal electoral manipulation when the citizen believes in the *SN* versus *BO*, this effect implies that manipulation is lower. However, there is another effect that encourages the leader to engage in higher levels of electoral manipulation. More manipulation implies that the joint probability of a weak leader and high economic performance is lower. This in turn leads the citizen to expect worse economic performance under a weak leader, reducing the expected utility of withholding support. This results in a force for complementarity between the *SN* and electoral manipulation.

We find that the former force dominates under some conditions, rendering manipulation and propaganda substitutes.

Our motivation in modeling the citizen's misperception in terms of causal processes is that causality is difficult to establish empirically, and often impossible with observational data. The *SN* is a plausible data generating process which produces credible estimates. To isolate the effect of causal misperceptions, we focus on a citizen is who otherwise rational, making optimal choices conditional on beliefs about the consequences of her actions. Given that data generating processes are typically unknowable without experimentation, the citizen displays a fairly mild behavioral bias that resonates with the complexity of the real world. Our contribution is to show that such misperceptions have major implications for the support anti-democratic politicians enjoy, the incentives to manipulate elections, 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 democratic backsliding (see e.g. Miller (2021); Howell et al. (2023); Horz (2021); Helmke, Kroeger and Paine (2022); Svolik (2020); 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, 2021). 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 (2020), we emphasize that expectations play a key role in explaining whether citizens support the incumbent, focusing on expectations about economic performance.

More generally, our theory contributes to research on autocratic rule. Scholars have scrutinized the role of elections (Little, 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, 2022).

Finally, 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 inaccurate beliefs (e.g., Minozzi, 2013; Bénabou and Tirole, 2006; Ashworth and Fowler, 2019; Little, Schnakenberg and Turner, 2022; Lockwood, 2017). Specifically, our paper builds on the framework proposed by Spiegler (2016), whose distinct focus is on causal relationships in a multivariate context. Other recent work also studies belief formation in a multivariate environment, but focus on heterogeneous preferences (Little, 2019) or ideological competition (Izzo, Martin and Callander, 2023). In our paper, citizens form beliefs about the causal relationship between political variables and economic outcomes. While scholars have investigated the role of the elites' beliefs on this relationship (Abramson and Montero, 2020; Albertus and Gay, 2017), formal work on the citizens' beliefs is less common.

Baseline Model

Setup Our model features infinitely many periods. In each period, a representative (or decisive) citizen chooses an action. All citizens and time periods are identical; hence, we do not introduce a time variable and simply refer to a representative citizen, C. The dynamic aspect of the model comes from the fact that the citizen can observe past choices and outcomes—what we refer to as the polity's history—and uses this information to form beliefs about the likelihood of outcomes in the current period.

The citizen chooses whether to support the incumbent, $a = 1$, or not, $a = 0$. For example, $a = 1$ could correspond to voting in favor of a referendum that would expand the autocrat's powers, supporting candidates in the incumbent's party, 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 $y = 1$.

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 the next section.[3] However, the citizen does care about economic performance, which may correspond to low unemployment, high growth, or low inflation. Her payoff is:

$$u_C = y + ax,$$

where x is a parameter that governs the citizen's relative ideological inclination to support the incumbent.[4] The citizen's ideology x is drawn from a twice-differentiable cumulative distribution function F with full support on $(-1, 1)$.

Consistent with theories emphasizing the ease with which incumbents can buy off potential challengers, we assume that the true data generating process is in the upper panel in Figure 1: citizen's support and economic performance jointly affect leader strength, but economic performance is exogenous and high with probability $\gamma \in (0, 1)$.

The DAGs in Figure 1 are entirely non-parametric and consistent with positive, negative, or non-monotonic effects. 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.

The *Strongman Narrative* is given in the lower panel 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

[3] As a result, the baseline model is also applicable to electoral politics in consolidated democracies.

[4] For example, suppose that the incumbent'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 the incumbent versus the challenger: $x = |x_I - x_V| - |x_C - x_V|$, where x_V is the voter's ideal point.

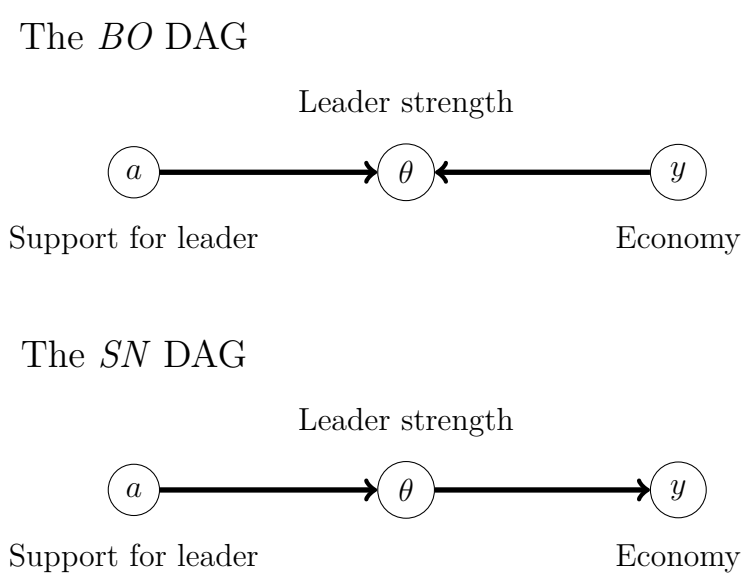

Figure 1: Two data generating processes linking citizen support (a), leader strength (θ), and economic performance (y): the *BO* DAG and the *SN* DAG.

Personal Equilibrium. The technical definition is in SM B; we focus on substantive considerations here. Citizens in our model can observe past outcomes, and because there are infinitely many periods, each 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 $\theta = ay$, the historical frequency of strong leaders 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
⋮	⋮	⋮	⋮
Marg. Prob.	β	γ	$\beta\gamma$

Given the information contained in the polity's history, the citizen forms beliefs about the effects of supporting the incumbent. Beliefs about *causal processes* matter because they prescribe how posteriors about the *variables* in the model should be formed (or "factored").[5] For example, when the citizen has rational expectations and believes in the

[5]Different from Izzo, Martin and Callander (2023) and Schwartzstein and Sunderam (2021), in our model citizens estimate parameters from the data that they observe, rather than being presented with a

BO DAG, the joint distribution of the variables a, θ, and y is:

$$\Pr_{BO}(a, \theta, y) = \Pr(a) \Pr(y) \Pr(\theta \mid a, y). \tag{1}$$

The joint distribution follows directly from the DAG: both support (a) and economic performance (y) are exogenous, while leader strength (θ) depends on both of these features. By contrast, the *SN* asserts the following joint distribution:

$$\Pr_{SN}(a, \theta, y) = \Pr(a) \Pr(\theta \mid a) \Pr(y \mid \theta). \tag{2}$$

All relevant beliefs can be deduced using standard probability operations from the joint distributions in Expressions 1 and 2.[6] Because conditional probabilities are generically not equal to the unconditional probabilities, different causal models lead to different beliefs about outcomes.

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). Indeed, 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 interests. This operation is critical to empirical inquiry. For example, regression is essentially an operation to approximate the conditional expectation function (Angrist and Pischke, 2008).

Once beliefs are formed, the citizen computes her subjective expected utility from each action and chooses an action accordingly. An important feature of our model is that the expected utility of an action today may depend on how often it was played in the past: the fraction of periods in the polity's history when a occurred. Furthermore, personal equilibrium requires that this long-run frequency be consistent with current behavior. Thus, we are looking for a number, $\beta^* \in [0, 1]$ that represents the probability with which the citizen supports the incumbent in each period. This probability needs to maximize

full model that already contains parameters.

[6] We use the notation Pr for objective probabilities and $\Pr_{BO}$ and $\Pr_{SN}$ for subjectively derived probabilities.

the citizen's expected utility and derived from the modified chain rule in Expressions 1 and 2, respectively.[7]

Discussion Before continuing with the analysis, we discuss several key features of the model. First, for clarity, while we refer to our variables as supporting the incumbent (a), leader strength (θ), and economic performance (y), we emphasize that other interpretations are also possible. For example, the citizen's action a can be volunteering for the regime/leader or not protesting. 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 also affect leader strength θ.

Second, we discuss how our approach relates to standard approaches. In most formal theories of politics, actors can properly anticipate the consequences of their actions because they know the data-generating process. For example, in models of political selection, protesting or voting against the leader is a gamble that can pay off when the new officeholder is better than the incumbent (Ashworth, 2012). Thus, while citizens face some uncertainty (i.e., they may have to form beliefs about specific states of the world), they do know the probability distribution over the consequences of their actions. By contrast, in our account citizens have misperceptions about the effects of their actions, because they form their beliefs using the modified 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 the (wrong) data generating process, etc.—to isolate the effect of causal misperceptions.

Third, we focus on a specific propaganda message, the *SN*, and study its implications for beliefs about the consequences of actions and behavior. This is consistent with empirical work that shows that some citizens may be convinced by pro-regime propaganda (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

[7]To 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 Expressions 1 and 2, respectively) and consistent with the citizen's strategy.

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 in which the receiver takes an action that is beneficial to the propagandist when the posterior belief regarding the relevant state of the world is sufficiently high. By contrast, while our decision stage is relatively straightforward as well, the receiver's decision-making is more complex because past behavior influences the incentives for current actions.

Analysis

Correct Expectations We start with a benchmark analysis of the model supposing that the representative citizen is fully aware of the true data generating process, namely that economic performance is exogenous, and that it—alongside her decision whether to support—determines leader strength. The joint distribution is given in Expression 1. The citizen cares about the marginal distribution of high economic performance, $y = 1$, conditional on her support decision, a. Because a and y are independent, the expectation is just γ—economic performance does not depend on support. Thus, 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. The citizen supports if $x \geq 0$ and the ex-ante probability of support is $1 - F(0)$.

Incorrect Expectations In contrast to the case of rational expectations, the joint distribution over the variables a, θ, and y, as factored by the propagandistic SN is the one derived in Expression 2. The citizen cares about the marginal distribution of y, conditional on the support choice a. This is equal to:

$$\Pr_{SN}(y \mid a) = \sum_{\theta} \Pr(\theta \mid a) \Pr(y \mid \theta). \tag{3}$$

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. In other words, 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, $\Pr(\theta \mid a)$, and the probability of economic performance conditional on leader strength, $\Pr(y \mid \theta)$.

Given *SN*, the expected utility of supporting the leader, $\Pr_{SN}(y = 1 \mid a = 1) + x$, is:[8]

$$\begin{aligned}
&\sum_{\theta} \Pr(\theta \mid a = 1) \Pr(y = 1 \mid \theta) + x \\
&= \Pr(\theta = 1 \mid a = 1) \Pr(y = 1 \mid \theta = 1) + \Pr(\theta = 0 \mid a = 1) \Pr(y = 1 \mid \theta = 0) + x \\
&= \gamma 1 + (1 - \gamma) \frac{\gamma(1 - \beta)}{1 - \beta\gamma} + x,
\end{aligned}$$

where, following Table 1, we denoted the endogenous, long-term frequency of citizen support with β. The first equality expands the sum, and the second equality plugs in the relevant numbers, as computed from historical joint probabilities. 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. Moreover, the citizen always expects high economic performance when the leader is strong, because, in fact, leader strength is *sufficient* for high economic performance: $\Pr(y = 1 \mid \theta = 1) = 1$. Second, the citizen expects high economic performance when the leader is weak with a lower probability which depends on past support: $\Pr(y = 1 \mid \theta = 0) = \frac{\gamma(1-\beta)}{1-\beta\gamma}$. This is a decreasing function of 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 illustrates this relationship for two different values of $\Pr(y = 1) = \gamma$.

[8] Detailed derivations can be found in the SM.

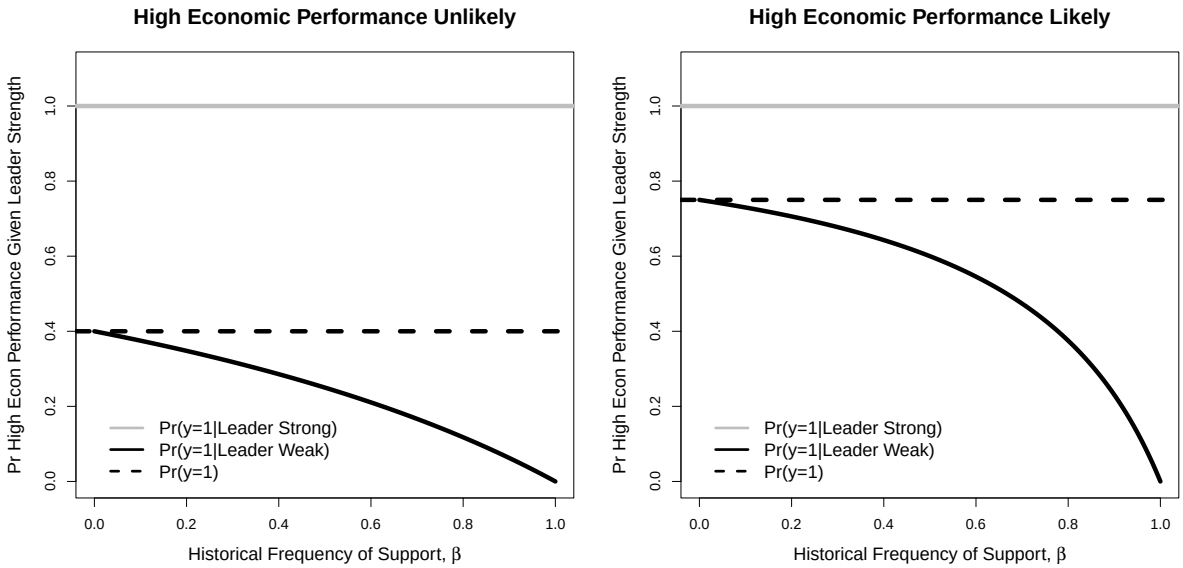

Figure 2: $\Pr(y = 1 \mid \theta = 1)$ and $\Pr(y = 1 \mid \theta = 0)$ as a function of β. Left panel: $\Pr(y = 1) = \gamma = 0.4$. Right panel: $\Pr(y = 1) = \gamma = 0.75$.

Again using Expression 3, the expected utility of not supporting is simply:

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

The intuition is that when the citizen does not support, the leader is always weak, and the citizen can expect high economic performance with probability $\frac{\gamma(1-\beta)}{1-\beta\gamma}$ as discussed above.

Here, 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 DGP.

It follows that a citizen who believes in the *SN* supports the incumbent if and only if:

$$\underbrace{x + \frac{\gamma(1-\gamma)}{1-\beta\gamma}}_{\text{NE}(\beta)} \geq 0. \tag{4}$$

The left-hand side represents the net expected utility of supporting, which is a function of long-run support (NE(β)). This is equal to the ideological benefit of supporting the incumbent plus the perceived increase in the probability of obtaining high economic performance when the citizen supports the leader.

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 incumbent on ideological grounds nevertheless support the incumbent because they (incorrectly) believe this may help obtain better economic outcomes. The second important observation from expression (4) 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.*

Intuitively, this is because 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 incumbent, 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 incumbent, leader strength and economic performance

are uncorrelated in the data, because the former is always weak regardless of economic performance.

Similar to cross-sectional complementarity in collective action—such as protests, strikes, or bank runs—historical complementarity means that each agent's participation leads to higher participation by others in equilibrium. But historical complementarity differs in two important ways. First, in our setting "other" agents exist in different periods. This kind of complementarity is asymmetric: past agents' participation influences those in the future, but not vice versa. Second, instead of payoff externalities that underlie standard, cross-sectional complementarity, historical complementarity works through a beliefs. When a polity's history features little support for the leader, the citizen today infers that the negative effect of withholding support on economic performance is small, *even when believing the regime's propaganda.* In contrast, if history features high support for the leader, the correlation between leader strength and economic performance is higher (due to reverse causality), and the citizen today estimates a higher effect of the former on the latter. Thus, the country's history teaches the citizen that the positive consequences of supporting the leader are higher, rendering support today more attractive.

The historical complementarity also has major implications for personal equilibria; for some values of ideological benefits x, multiple equilibria exist:

Proposition 1. *There exist thresholds $x_0 \equiv -\gamma$ and $x_1 \equiv -\gamma(1-\gamma)$, such that there is a unique no support equilibrium ($\beta^* = 0$) if $x < x_0$, and a unique always support equilibrium ($\beta^* = 1$) if $x > x_1$. When $x \in [x_0, x_1]$, both of these exists, along with a mixed strategy equilibrium ($\beta^* \in (0,1)$).*

Figure 3 illustrates the baseline analysis. Intuitively, for high levels of ideological payoffs, even if there were 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, the citizen never supports, because even concerns about the economy cannot overcome the citizen's strong ideological dislike of the incumbent. Finally, in the intermediate region, multiple personal equilibria exist that differ in the likelihood of supporting

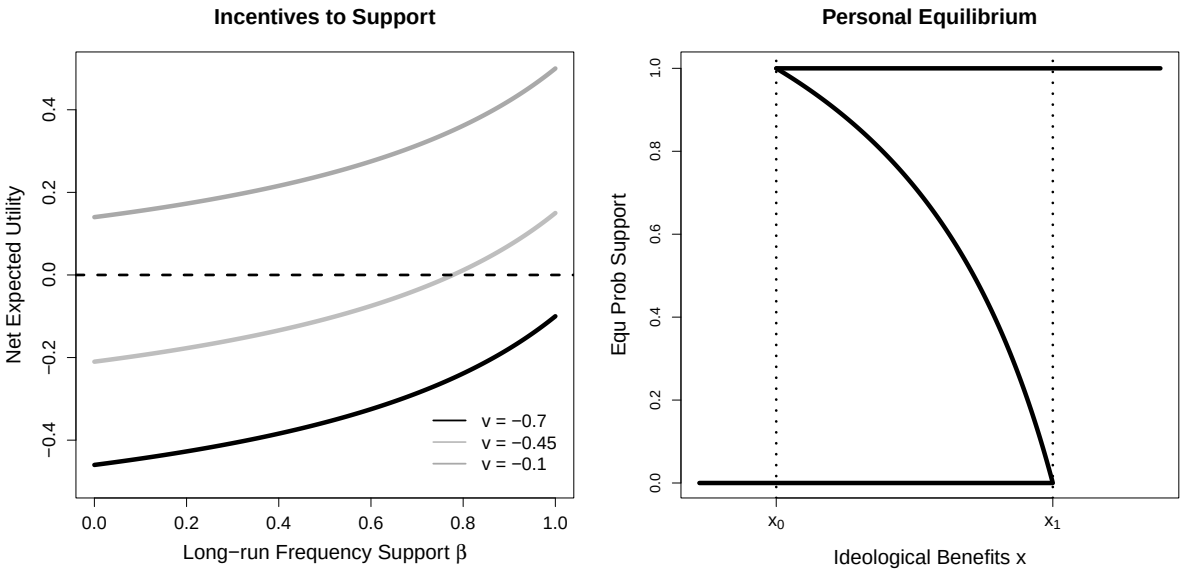

Figure 3: The Citizen's net expected utility (left panel) and the personal equilibria of the model (right panel). Parameter values: $\gamma = 0.6$.

the incumbent. Here, 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 optimal. Besides the never-support and always-support equilibria, there is also an equilibrium in which the citizen sometimes supports and sometimes does not. It is worth mentioning, however, that this interior equilibrium is knife-edge—a small perturbation would cause citizens to estimate a strictly positive or strictly negative net expected utility, resulting in a switch to one of the pure strategy equilibria. As a result, we focus on the pure strategy personal equilibria for the remainder of the paper.

An important implication of the analysis is that polities with relatively similar conditions (in terms of ideological affinities to the leader and economic performances) can be very different in terms of their experience of autocratic leader support—they form distinct "authoritarian legacies." Besides differences in citizens' utility functions, this effect contributes to our understanding of why similarly situated countries differ in terms of how much popular support their autocrats enjoy.

Next, we examine the effectiveness of the *SN* by comparing support under the *BO* and the *SN*. It is clear from Expression (4) that the leader always enjoys more support under the *SN*. To quantify the benefit of the *SN* for the leader, we examine each equilib-

rium separately. Because the mixed strategy equilibrium is not stable, we focus on pure strategy equilibria here. Using the results derived above, the effectiveness of the *SN* is:

$$\text{Eff}^{SN} \equiv 1 - F\left(x^{SN}\right) - \left[1 - F\left(x^{BO}\right)\right],$$

where $x^{BO} = 0$, and $x^{SN} \in \{x_0, x_1\}$ are the thresholds identified in Proposition 1. We can show the following:

Proposition 2. *When the always-support equilibrium is selected, a higher* γ *increases the effectiveness of the SN. When the never-support equilibrium is selected, the effectiveness of the SN is maximized at* $\gamma = 0.5$ *and monotonically decreases as* γ *is further from* 0.5*.*

The intuition is as follows. When the always-support equilibrium is selected, we say that the country has an authoritarian legacy. 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 and a weak inferred effect does little to encourage her. As γ increases, the citizen believes her support is increasingly effective in ensuring leader strength. Moreover, given that the citizen believes in the *SN*, the probability of high economic performance is very high when the leader is strong (formally, $p(y = 1 \mid \theta = 1) = 1$). This helps the *SN* to elicit support from the citizen, which makes always-support easier to sustain. Thus, in polities with an authoritarian legacy, we expect leaders to be able to enhance their powers in periods of economic prosperity.

When the never-support equilibrium is selected, we say that the country does not have an authoritarian legacy. In this situation, 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 probability of high economic performance also makes it more likely to the citizen expects good economic performance when the leader is weak, which makes not supporting a more attractive option, hurting the effectiveness of the *SN*. This latter effect is stronger when $\gamma > 0.5$. It follows that in countries without an authoritarian legacy, the *SN* is most effective in periods of economic

volatility—when γ is close to 0.5.

Computing Causal Effects and Statistical Sophistication It is useful to relate the citizen's perception to alternative approaches to forming beliefs about the effect of support on economic performance. A natural way to approach this issue would be to compute an *unconditional* difference-in-means estimate of the effect of supporting the incumbent on economic performance, which would yield the null effect as described by the *BO* DAG. A citizen who believes in the *SN* departs from a simple difference-in-means estimator in two ways:

1. The citizen believes that a is the only factor that affects θ: a treatment effect of support on the mediator variable leader strength.
2. The citizen believes that θ causes y and therefore conditions her expectation of economic performance on the post-treatment variable leader strength.

Because of these differences, the citizen's estimator yields a different conclusion.

The above discussion shows that the citizen makes a key mistake when considering the effect of support on economic performance: *conditioning on a collider*. In this respect, the citizen commits the same error social scientists often do—even those well-trained in empirical methods (Montgomery, Nyhan and Torres, 2018). Current practice in social science promotes using a range of estimators, and recent research finds that populist strongman generally lead to worse economic performance (Funke, Schularick and Trebesch, 2023). The citizen in our model, however, is less statistically sophisticated, relying on a single estimator.

Endogenous Beliefs and Propaganda As discussed, the citizen's beliefs *about data generating processes* are exogenous: the citizen either believes that the data was generated by the *SN* or the *BO*. But the citizen's beliefs *about the consequences of her actions* are endogenous, because they are derived from the data using beliefs about the causal process. We argue that the latter are "deeper" than the former—less amenable to falsification and more complex than beliefs about "states of the world." A natural question is where beliefs

about data generating processes come from. One explanation is that these are transmitted via socialization or cultural institutions, as in Weber's traditional legitimacy (Weber, 2004). Another explanation is that they are created or reinforced by propaganda. This process may work quite differently than standard communication games (for a review, see Little, 2023), because the citizens have to be convinced of a *causal network*, rather than the realization of a single variable. While fully endogenizing the citizen's causal model is beyond the scope of the present article, in the SM we present a leader's incentive to exert effort to convince the citizen that the *SN* is the true data generating process. We show that the leader's optimal "persuasion effort" is ambiguously related to the level of economic performance, showing that leaders in all kinds of economic environment may rely on the *SN* to elicit support.

Incorporating Democratic Backsliding

The baseline model is designed to present the *SN* in its simplest form. In this section, we enrich the baseline model to feature two notions of democratic backsliding: electoral manipulation and executive aggrandizement. We assume that the former is observed by the citizen while the latter occurs with some probability if the leader stays in power.

Formally, besides the existing variables of support a, leader strength θ, and economic performance y, we introduce the following additional variables. We denote by $m \in [0, 1]$ the level of electoral manipulation, and by $r \in \{0, 1\}$ whether the incumbent stays in power ($r = 1$). Electoral manipulation improves the chances of the incumbent retaining power:

$$\Pr(r = 1) = a + (1 - a)m.$$

This means that with probability m, the incumbent stays in power even if the citizen does not support him ($a = 0$). Furthermore, staying in power is a necessary condition for being a strong leader, with the other cause being high economic performance: $\theta = ry$. This is a natural generalization of our baseline specification (where $\theta = ay$). Implicit here is the assumption that if the incumbent is replaced, the challenger cannot be strong—i.e.

has no plans to increase his or her power.

To study concerns about executive aggrandizement, we introduce two more variables: $t \in \{0, 1\}$ and $b \in \{0, 1\}$. The former indicates whether or not there is an opportunity to engage in executive aggrandizement: $t = 1$ with probability $q \in (0, 1)$ and we say that the leader has an opportunity. The latter indicates whether or not aggrandizement occurs, which happens if and only if the leader is strong and the opportunity is present: $b = \theta t$. Thus, we distinguish between (institutional or more generally political) leader strength and executive aggrandizement. This distinction is motivated substantively: institutions that give more power to the current officeholder do not necessarily represent a degradation in democratic quality. The citizen understands this difference. She also knows that electoral manipulation can keep the incumbent in office. But, similar to the baseline model, she may have incorrect beliefs about the data generating process. Specifically, we assume that as before, the *SN* inverts the link from y to θ—see the DAGs in Figure 4.

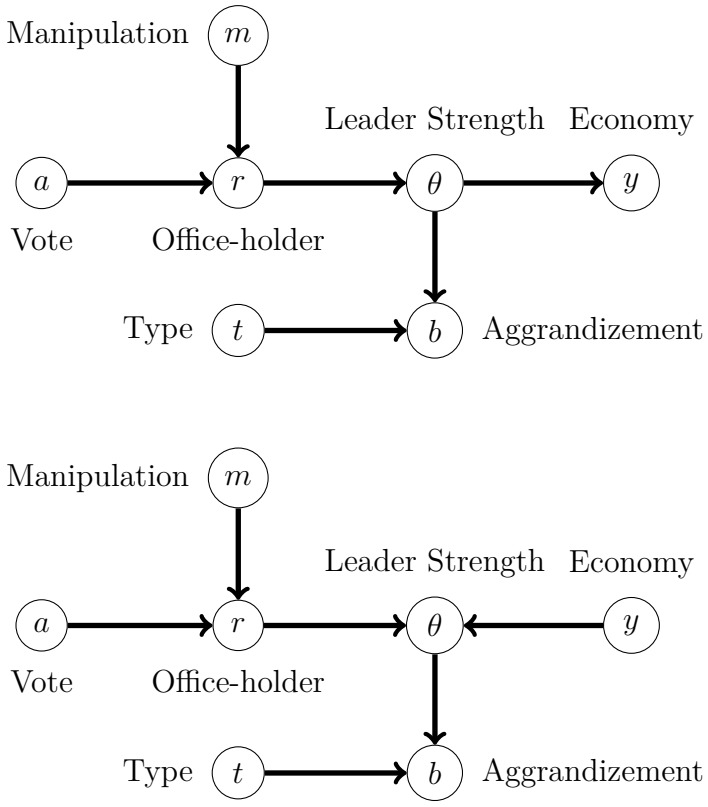

Figure 4: The *BO* and the *SN* DAG when backsliding can occur.

The citizen's payoff is now given by:

$$u_C = y + rx - b\psi,$$

where $\psi \geq 0$ denotes the citizen's distaste for democratic backsliding. Moreover, compared to the baseline model, the net ideological payoff x is now obtained if the incumbent remains in office, $r = 1$, rather than purely as a function of the citizen's support decision.

Correct Expectations We begin by analyzing a citizen who has the *BO* DAG. The expected payoff of supporting the leader is $\gamma + x - \psi q\gamma$, and the expected payoff of withholding support is $\gamma + m(x - \psi q\gamma)$. The intuition is that the citizen knows that y is exogenous, and equal to 1 with probability γ. The citizen receives the net ideological payoff x for sure when she supports, and with probability m otherwise. Similarly, aggrandizement is a more salient concern if she supports: it can happen when there is both opportunity for it and the economy is good. Without the citizen's support, besides these factors, electoral manipulation also has to be successful to culminate in executive aggrandizement. Re-arranging, the citizen supports if

$$x \geq \gamma q \psi. \tag{5}$$

Note that from an ex-ante perspective, the probability of support here is $1 - F(\gamma q \psi)$. This is a higher threshold than the baseline model, driven by concerns about executive aggrandizement. Importantly, this threshold is independent of electoral manipulation m.

Incorrect Expectations Here, the citizen needs to compute the marginal distributions of y, b, and r. The variable r is not altered by the *SN*, so the citizen gets x if she supports and mx otherwise, as before. Consider y next. According to the *SN*, the marginal distribution of $y = 1$ is given by:

$$\Pr_{SN}(y = 1 \mid a, m) = \sum_{r,\theta,b,t} \Pr(t) \Pr(r \mid a, m) \Pr(\theta \mid r) \Pr(y = 1 \mid \theta) \Pr(b \mid t, \theta).$$

The expression is conditional on support a and manipulation m because the citizen chooses a and observes m; hence, the citizen can condition on these variables.

Suppose that the citizen supports the incumbent, which implies he remains in office. The key terms then are $\Pr(\theta \mid r)$ and $\Pr(y = 1 \mid \theta)$. When $r = 1$, we have $\Pr(\theta = 1 \mid r = 1) = \gamma$ and $\Pr(\theta = 0 \mid r = 1) = 1 - \gamma$. The citizen next needs to calculate the expectation of economic performance conditional on leader strength: $\Pr(y = 1 \mid \theta = 0)$ and $\Pr(y = 1 \mid \theta = 1)$. Similar to the baseline case, we have $\Pr(y = 1 \mid \theta = 1) = 1$ because $y = 1$ is a *necessary* condition for a strong leader. Moreover,

$$\Pr(y = 1 \mid \theta = 0) = \frac{\Pr(r = 0)\gamma}{1 - \Pr(r = 1)\gamma},$$

where β is again the long-run frequency of support and $\Pr(r = 1) = \beta + (1 - \beta)m$.

Plugging in, $\Pr_{SN}(y = 1 \mid a = 1, m)$ is equal to:

$$\gamma \Pr(y = 1 \mid \theta = 1) + (1 - \gamma) \Pr(y = 1 \mid \theta = 0) = \gamma + (1 - \gamma) \left(1 - \frac{1 - \gamma}{1 - (\beta + (1 - \beta)m)\gamma}\right).$$

Moreover, the citizen has to figure out the marginal probability of aggrandizement, i.e., $b = 1$. This is simply equal to γq, since there is only aggrandizement if both $\theta = 1$ and $t = 1$. This is the same as with rational expectations, because the *SN* does not alter the causal mapping between these variables.

Putting everything together, the expected utility of support is:

$$\gamma + (1 - \gamma) \left(1 - \frac{1 - \gamma}{1 - (\beta + (1 - \beta)m)\gamma}\right) - \psi\gamma q + x.$$

Now suppose that the citizen does not support, which renders both $r = 1$ and $r = 0$ possible. If $r = 1$ (which happens with probability m), both a strong ($\theta = 1$) and a weak leader ($\theta = 0$) are possible again. Otherwise, only a weak leader ($\theta = 0$) is possible. The key expressions $\Pr(\theta \mid r)$ and $\Pr(y = 1 \mid \theta)$ remain unchanged. Hence, the perceived probability of high economic performance when the citizen withholds support,

$\Pr_{SN}(y = 1 \mid a = 0)$, is:

$$m\left[\gamma \Pr(y = 1 \mid \theta = 1) + (1 - \gamma) \Pr(y = 1 \mid \theta = 0)\right] + (1 - m) \Pr(y = 1 \mid \theta = 0).$$

Moreover, the marginal with respect to backsliding is simply $m\gamma q$, as in case where the citizen holds correct expectations. Therefore, the expected utility of not supporting is equal to:

$$m\left[\gamma \Pr(y = 1 \mid \theta = 1) + (1 - \gamma) \Pr(y = 1 \mid \theta = 0) + x - \psi\gamma q\right] + (1 - m) \Pr(y = 1 \mid \theta = 0).$$

Re-arranging and simplifying, the citizen supports if

$$\underbrace{x + \gamma\left[1 - \Pr(y = 1 \mid \theta = 0)\right] - \gamma\psi q}_{\equiv \mathrm{NE}(\beta)} \geq 0.$$

Comparing this with the correct expectations case, the left-hand side is larger. Hence, best response support for any level of m is weakly larger: the *SN* helps the leader enjoy more support.

Recall that $\Pr(y = 1 \mid \theta = 0) = 1 - \frac{1-\gamma}{1-(\beta+(1-\beta)m)\gamma}$. This expression is decreasing in β and m. As a result, the net expected utility of support, $\mathrm{NE}(\beta)$ is *increasing* in β. Thus, similar to the benchmark analysis, the game features historical complementarities, which can result in multiple personal equilibria:

Proposition 3. *There exists thresholds* $\tilde{x}_0 \equiv \gamma(\psi q - 1)$ *and* $\tilde{x}_1 \equiv \gamma\left(\psi q - \frac{1-\gamma}{1-m\gamma}\right)$*, such that if* $x < \tilde{x}_0$*, there is a unique never-support equilibrium (*$\beta^* = 0$*), and if* $x > \tilde{x}_1$*, there is a unique always-support equilibrium (*$\beta^* = 1$*). When* $x \in [\tilde{x}_0, \tilde{x}_1]$*, both of these exists, along with a mixed strategy equilibrium (*$\beta^* \in (0, 1)$*).*

Figure 5 illustrates the equilibria, showing that the citizen's net expected utility is increasing in the long-run frequency of support (left panel) and that for some parameter values, multiple equilibria exist (right panel).

Similar to the baseline case, it is instructive to calculate the effectiveness of the *SN*.

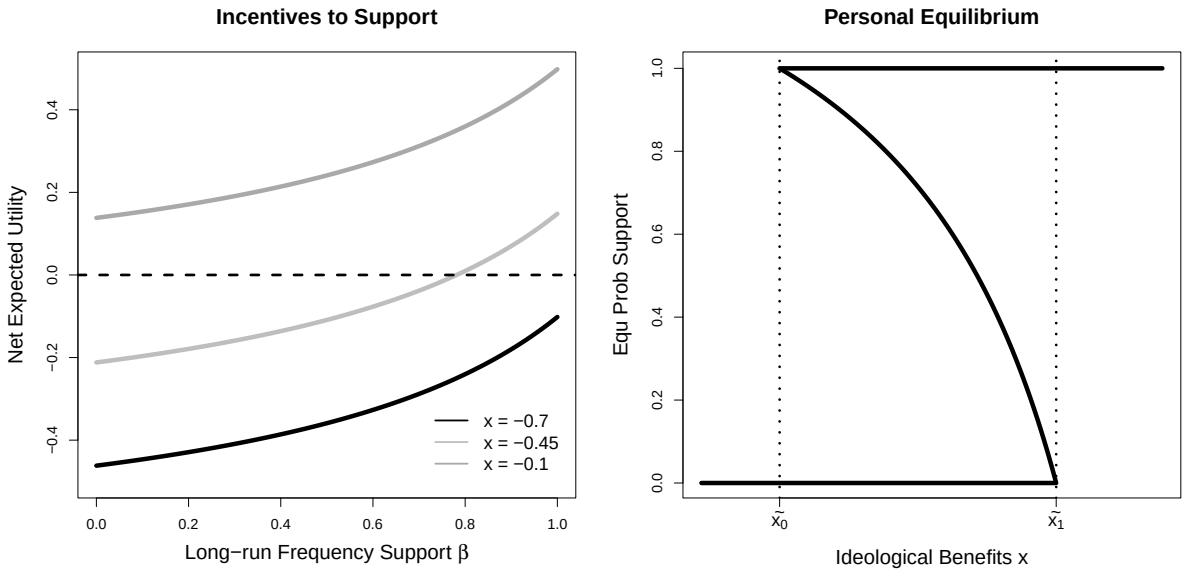

Figure 5: The Citizen's net expected utility (left panel) and the personal equilibria of the model (right panel). Parameter values: $\gamma = 0.6$, $\psi = 0.01$, $q = 0.3$, and $m = 0$.

In a pure strategy equilibrium, the effectiveness of the *SN* is given by:

$$\text{Eff}^{SN} = 1 - F\left(\tilde{x}^{SN}\right) - \left[1 - F\left(\tilde{x}^{BO}\right)\right].$$

where $\tilde{x}^{BO} \equiv \psi q \gamma$ and $\tilde{x}^{SN} \in \{\tilde{x}_0, \tilde{x}_1\}$ are identified in Proposition 3. The effectiveness of the *SN* depends on the probability of obtaining high economic performance, γ, and on the level of electoral manipulation, m. Similar to Proposition 2, one can show that the effect of γ on Eff^{SN} is ambiguous.[9] We now examine how the effectiveness of the *SN* changes when there is a higher level of electoral manipulation:

Proposition 4. *In a pure strategy equilibrium, a higher level of electoral manipulation increases the effectiveness of the SN.*

Intuitively, when there is more electoral manipulation, the probability of obtaining high economic performance conditional on a weak leader, $\Pr(y = 1 \mid \theta = 0)$, decreases. This is because, similar to the effect of higher β, the joint probability of high economic performance and a weak leader falls as leaders tend to be strong more often. This makes withholding support less attractive, increasing the effectiveness of the *SN*.

[9] In fact, this is more complex because the threshold $\psi q \gamma$ that is utilized when the citizen has rational expectations now also depends on γ. See the SM for further details.

Endogenous Electoral Manipulation

We now allow the incumbent to endogenously choose the optimal level of electoral manipulation m. The incumbent's utility function is:

$$u_I = r + \omega_\theta \theta + \omega_b b - k(m),$$

where $\omega_\theta \geq 0$ is concern for greater power, $\omega_b \geq 0$ is concern for executive aggrandizement, and k is the cost function for electoral manipulation. In other words, the incumbent wishes to stay in power and expand his powers, up to the point for engaging in executive aggrandizement. Manipulation is costly, with a cost function k that is increasing and convex.

Finally, the incumbent does not know the true level of x, but knows it is drawn from F. The sequence of the game is:

1. Incumbent chooses manipulation $m \in [0, 1]$.
2. The ideological affinity of the representative citizen x is drawn from F.
3. Citizen observes x and m and chooses $a \in \{0, 1\}$.

We look for a tuple (β^*, m^*) such that β^* is a personal equilibrium (given the realized value of x and any $m \in [0, 1]$) and m^* maximizes the incumbent's payoff function, given the anticipated probability of support. We consider the cases where the citizen has correct and incorrect expectations, and compare the optimal levels of manipulation. The incumbent knows the true data generating process and whether or not the citizen believes in the *SN*.

Correct Expectations Consider first the case in which the citizen has rational expectations. By Expression 5, the probability that the citizen supports is:

$$1 - F\left(\tilde{x}^{BO}\right) = 1 - F\left(\psi q \gamma\right).$$

This quantity does not vary with m. As a result, the incumbent solves:

$$\max_{m\in[0,1]} \left[1 - F\left(\tilde{x}^{BO}\right) + F\left(\tilde{x}^{BO}\right) m\right] (1 + \omega_\theta \gamma + \omega_b \gamma q) - k(m)$$

The intuition is that the probability that the incumbent remains in office ($r = 1$) is $1-F(\psi q\gamma)+F(\psi q\gamma)\, m$. When in office, the leader is strong with probability γ (and gets a payoff of ω_θ). Moreover, when the leader is strong and the opportunity for aggrandizement presents itself, the incumbent receives an additional payoff of ω_b.

The associated first-order condition is:

$$F\left(\tilde{x}^{BO}\right) (1 + \omega_\theta \gamma + \omega_b \gamma q) = k'(m).$$

Intuitively, the lower the probability the citizen supports the incumbent ($F\left(\tilde{x}^{BO}\right) = F(\psi q\gamma)$), the higher the optimal level of manipulation.

Incorrect Expectations Now consider the case where the citizen believes in the *SN*. As detailed above, in a pure strategy equilibrium, the probability of supporting the leader is: $1 - F(\tilde{x}^{SN})$, where $\tilde{x}^{SN} \in \{\tilde{x}_0, \tilde{x}_1\}$ is one of the equilibrium thresholds identified in Proposition 3, depending on the historical legacy. Then, the incumbent solves:

$$\max_{m\in[0,1]} \left[1 - F\left(\tilde{x}^{SN}\right) + F\left(\tilde{x}^{SN}\right) m\right] (1 + \omega_\theta \gamma + \omega_b \gamma q) - k(m).$$

The first-order condition is:

$$\left(\underbrace{-f\left(\tilde{x}^{SN}\right) \frac{\partial \tilde{x}^{SN}}{\partial m}(1-m)}_{\text{Change in Support}} + \underbrace{F\left(\tilde{x}^{SN}\right)}_{\text{No-Support}} \right) (1 + \omega_\theta \gamma + \omega_b \gamma q) = k'(m).$$

From Proposition 4, $\frac{\partial \tilde{x}^{SN}}{\partial m} \leq 0$. Hence, the expression labeled "Change in Support" is positive, pushing the incumbent to choose a higher level of electoral manipulation. The reason is that a higher level of electoral manipulation makes the citizen less confident in obtaining high economic performance when the leader is weak.

However, the probability of *not* supporting the leader, $F(\tilde{x}^{SN})$, is smaller than the corresponding probability under rational expectations: $F(\psi q \gamma)$. This effect pushes the incumbent to choose a lower level of m, compared to the case in which the citizen has rational expectations. Which of these effects dominate, and thus whether equilibrium manipulation is higher under the *SN*, depends on the precise functional form of F. In SM D we let F be given by the Uniform distribution on $(-1, 1)$ and show that optimal manipulation is *lower* when the citizen believes in the *SN*, regardless of equilibrium selection. Hence, in this case, electoral manipulation and propaganda act as substitutes.

Returning to the general case in which F is arbitrary, the incumbent is always better off when the citizen believes in the *SN*:

Proposition 5. *The incumbent's equilibrium utility is higher when the citizen believes in the SN.*

This provides a justification for investing in the *SN*. Importantly, this result holds regardless of the which (pure strategy) personal equilibrium is played by the citizen.

This result is illustrated in Figure 6. Here, the optimal level of manipulation when the citizen believes in the *SN* can take different forms, depending on parameter values. In the left panel, high economic performance is rare, and so the difference in optimal manipulation levels is low. By contrast, in the right panel, high economic performance occurs with intermediate frequency, and so the difference in optimal manipulation levels is larger. Nevertheless, it is always the case that the incumbent's equilibrium utility is higher when the citizen believes in the *SN*, as shown in Proposition 5.

Conclusion

In their propaganda campaigns, autocrats and would-be autocrats around the world emphasize the benefits of strong states via a *Strongman Narrative*. What are the correlates and the behavioral implications of believing these propaganda messages compared to having rational expectations? Analyzing a formal model of belief formation and support behavior, we emphasize that the effectiveness of the *SN* depends on history, ideological

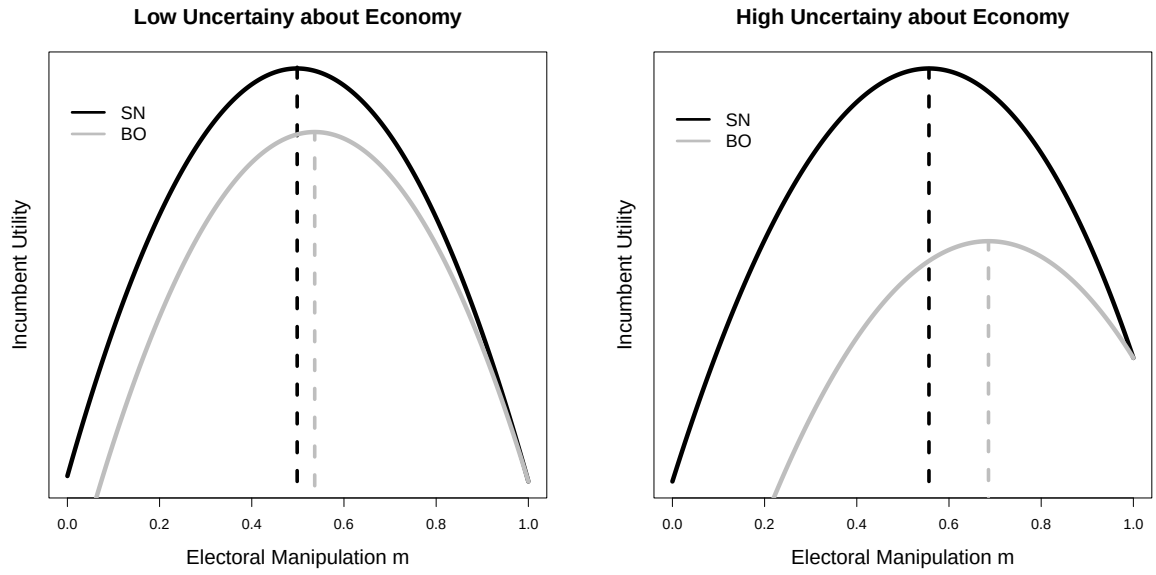

Figure 6: The incumbent's objective function and optimal levels of electoral manipulation when the citizen believes in the *SN* (black lines) and when the citizen believes in the *BO* DAG (gray lines). Parameter values, both panels: $\psi = 0.1$, $q = 0.4$, $\omega_\theta = 0.7$, $\omega_b = 0$, and $k(m) = \frac{1}{2}m^2$. Left panel: $\gamma = 0.1$. Right panel: $\gamma = 0.5$.

considerations, and electoral manipulation. In particular, having supported a strongman before increases the incentives to support a strongman today, which is a form of historical complementarity that results in multiple (personal) equilibria for intermediate levels of net ideological payoffs (i.e., for moderates). Moreover, electoral manipulation increases the effectiveness of the *SN*, which pushes incumbents to choose a higher level of it. However, precisely because the *SN* increases support, it also makes electoral manipulation less necessary. Hence, citizens believing in the *SN* has competing effects on the optimal level of electoral manipulation, while always increasing the incumbent's welfare.

In our model, we abstracted away from a number of other important features in the process of democratic backsliding. In particular, our model does not give the challenger—and more generally other political actors besides the incumbent and a representative voter—an active role. In the real world, these other political actors are important for affecting outcomes, often playing the role of a "vertical restrainer" (Grillo et al., 2024). Future work might investigate to what extent the effectiveness of the challenger (or a court) as a restrainer is affected by the citizen believing in the *SN* or not.

While we have focused on a commonly employed propaganda message, the *SN*, future

work should expand the analysis to scrutinize the behavioral implications of different propaganda claims. For example, a commonly employed strategy is to blame bad outcome on minorities or foreign powers. Such propaganda claims typically co-exist with propaganda claims regarding the importance of strong leaders like the *SN*. Analyzing the interaction between these propaganda claims can be an important avenue for future work.

References

Abramson, Scott F and Sergio Montero. 2020. "Learning about Growth and Democracy." *American Political Science Review* 114(4):1195–1212.

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: Princeton university press.

Ashworth, Scott. 2012. "Electoral Accountability: Recent Theoretical and Empirical Work." *Annual Review of Political Science* 15:183–201.

Ashworth, Scott and Anthony Fowler. 2019. "Electorates vs. Voters." *Unpublished Manuscript. University of Chicago* .

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*. 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. 2021. "A Wolf in Sheep's Clothing? Citizen Uncertainty and Democratic Backsliding." *Paper presented at the APSA Annual Meeting* .

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.

Fearon, James D. 2011. "Self-enforcing Democracy." *The Quarterly Journal of Economics* 126(4):1661–1708.

Frantz, Erica. 2018. *Authoritarianism: What Everyone Needs to Know®*. Oxford University Press.

Funke, Manuel, Moritz Schularick and Christoph Trebesch. 2023. "Populist leaders and the economy." *American Economic Review* 113(12):3249–3288.

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.

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. 2020. "Reference points and democratic backsliding." *American Journal of Political Science* .

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*. Broadway Books.

Little, Andrew T. 2012. "Elections, Fraud, and Election Monitoring in the Shadow of Revolution." *Quarterly Journal of Political Science* 7(3):249–283.

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. 2023. "Bayesian explanations for persuasion." *Journal of Theoretical Politics* 35(3):147–181.

Little, Andrew T, Keith E Schnakenberg and Ian R Turner. 2022. "Motivated reasoning and democratic accountability." *American Political Science Review* 116(2):751–767.

Lockwood, Ben. 2017. "Confirmation bias and electoral accountability.".

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.

Montgomery, Jacob M., Brendan Nyhan and Michelle Torres. 2018. "How conditioning on posttreatment variables can ruin your experiment and what to do about it." *American Journal of Political Science* 62(3):760–775.

Morgan, Stephen L. and Christopher Winship. 2015. *Counterfactuals and Causal Inference*. Cambridge: Cambridge University Press.

Öztürk, Aykut. 2022. "Whisper Sweet Nothings to Me Erdogan: How Economic Propaganda Works Under Authoritarianism.".

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.

Schwartzstein, Joshua and Adi Sunderam. 2021. "Using Models to Persuade." *American Economic Review* 111(1):276–323.

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.

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*. Hackett Publishing.

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

Contents

A Proofs

Proof of Proposition 1. Using the definition of the net expected utility of supporting, there is a unique always-support personal equilibrium if

$$\text{NE}(\beta = 0) > 0 \Rightarrow x > -\gamma(1-\gamma) \equiv x_1.$$

Moreover, there is a unique never-support equilibrium if

$$\text{NE}(\beta = 1) < 0 \Rightarrow x < -\gamma \equiv x_0.$$

Observe that $x_0 < x_1$.

Finally, we have that there is a mixed strategy personal equilibrium if

$$\text{NE}(\beta^I) = 0 \Rightarrow \beta^I = \frac{-x - \gamma(1-\gamma)}{\gamma(-x)}.$$

This is interior if $x > x_0$ and $x < x_1$. □

Proof of Proposition 2. Recall that the effectiveness of the *SN* is given by:

$$\text{Eff}^{SN} \equiv 1 - F\left(x^{SN}\right) - \left[1 - F\left(x^{BO}\right)\right],$$

where $x^{BO} = 0$ and $x^{SN} \in \{x_0, x_1\}$ are the thresholds identified in Proposition 1. Suppose first that $x_0 = -\gamma$ is the equilibrium threshold. Then:

$$\frac{\partial \text{Eff}^{SN}}{\partial \gamma} = -f(x^{SN})(-1) > 0.$$

So an increase in γ increases the effectiveness of the *SN*.

Now suppose that $x_1 = -\gamma(1-\gamma)$ is the equilibrium threshold. Then:

$$\frac{\partial \text{Eff}^{SN}}{\partial \gamma} = -f(x^{SN})(-1+2\gamma) > 0.$$

So an increase in γ increases the effectiveness of the *SN* if $\gamma < \frac{1}{2}$ and decreases it otherwise. □

Proof of Proposition 3. Using the definition of the net expected utility of supporting, there is a unique always-support personal equilibrium if

$$\text{NE}(\beta = 0) > 0 \Rightarrow x > \gamma \left(\psi q - \frac{1-\gamma}{1-m\gamma} \right) \equiv \tilde{x}_1.$$

Moreover, there is a unique never-support equilibrium if

$$\text{NE}(\beta = 1) < 0 \Rightarrow x < \gamma \left(\psi q - 1 \right) \equiv \tilde{x}_0.$$

Observe that $v_0 < v_1$.

Finally, we have that there is potentially a mixed strategy personal equilibrium if

$$\text{NE}(\beta^I) = 0 \Rightarrow \beta^I = \frac{1}{\gamma(1-m)} \left[1 - m\gamma - \frac{\gamma(1-\gamma)}{\psi\gamma q - x} \right]$$

By inspection, β^I is decreasing in x and decreasing in m. It is interior if $x \in (\tilde{x}_0, \tilde{x}_1)$. □

Proof of Proposition 4. Recall the definition of effectiveness:

$$\text{Eff}^{SN} = 1 - F\left(\tilde{x}^{SN}\right) - \left[1 - F\left(\tilde{x}^{BO}\right)\right],$$

where $\tilde{x}^{BO} = \psi q \gamma$ and $\tilde{x}^{SN} \in \{\tilde{x}_0, \tilde{x}_1\}$ are defined in Proposition 3. Observe that $\tilde{x}_0$ is not a function of m while

$$\frac{\partial \tilde{x}_1}{\partial m} = -\gamma(1-\gamma)(-1)(1-m\gamma)^{-2}(-\gamma) = \frac{-\gamma^2(1-\gamma)}{(1-m\gamma)^2} < 0.$$

As a consequence:

$$\frac{\partial \text{Eff}^{SN}}{\partial m} = -f\left(\tilde{x}^{SN}\right) \frac{\partial \tilde{x}^{SN}}{\partial m} \geq 0.$$

Hence, an increase in m weakly increases the effectiveness of the *SN*. □

Proof of Proposition 5. Define $w \equiv 1 + \omega_\theta \gamma + \omega_b \gamma q$. Further define $\Pr(\text{win} \mid m, BO) \equiv$

$1 - F\left(\tilde{x}^{BO}\right) + F\left(\tilde{x}^{BO}\right) m$ and $\Pr\left(\text{win} \mid m, SN\right) \equiv 1 - F\left(\tilde{x}^{SN}\right) + F\left(\tilde{x}^{SN}\right) m$ as the probabilities of staying in power given m when the citizen believes in the BO and SN DAGs respectively.

Using this notation, the incumbent's expected payoff when the citizen believes in the BO can be written as:

$$\Pr\left(\text{win} \mid m, BO\right) w - k(m)$$

Similarly, the incumbent's payoff when the citizen believes in the SN is:

$$\Pr\left(\text{win} \mid m, SN\right) w - k(m)$$

First, observe that *for a fixed level of electoral manipulation*, m, the probability of winning is higher when the citizen believes in the SN:

$$\Pr\left(\text{win} \mid m, SN\right) > \Pr\left(\text{win} \mid m, BO\right)$$

This is because the popular support of the citizen is higher when the SN is employed as the causal map guiding the citizen's decision-making. Given that the remaining portions of the utility functions are the same (w and $k(m)$), the incumbent's payoff is higher for any fixed level of m.

Second, note that the incumbent chooses m to maximize the probability of winning minus the costs of manipulation. When the citizen believes in the SN, the incumbent *can* choose the same level of manipulation as when the citizen believes in the BO DAG, and when doing so, receives a higher utility. When a different level of manipulation is chosen, it must be the case that the probability of winning is even higher for this alternative level of manipulation, *at least* compensating for higher costs of electoral manipulation. Hence, the incumbent's equilibrium utility is weakly higher when the citizen believes in the SN. □

B Technical Details on Equilibrium Concept

In this section, we briefly review the approach laid out in Spiegler (2016). Let $x = (x_i)_{i=1,\dots,n}$ be the collection of variables under consideration and $p(x)$ its joint distribution. Given a DAG R, the subjective joint probability of a set of events $p_R(x)$ is calculated by multiplying the probabilities of each event conditional on their causes:

$$p_R(x) = \prod_{i=1}^{n} p(x_i \mid x_{R(i)}), \tag{6}$$

where $R(i)$ is the set of direct parents of the node i.[10] From the joint distribution $p_R(x)$ all relevant beliefs can be deduced using the usual probability operations. Because the probability $p(x_i)$ is generically not equal to the probability $p(x_i \mid x_j)$, different causal models may lead to different inferences.[11]

Once beliefs are formed, the citizen computes her expected utility for each action and chooses the action that promises the highest level of expected utility. 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 and let y be a payoff-relevant variable. A distribution $p \in \Delta(x)$ with full support on the choice set A is an ϵ-perturbed personal equilibrium if*

$$a \in \arg\max_{a'} \sum_{y} p_R(y \mid a) u(a', y)$$

whenever $p(a) > \epsilon$. A distribution p^* is a personal equilibrium if there exists a sequence $p^k \to p^*$ of perturbations of p^*, as well as a sequence $\epsilon^k \to 0$, such that p^k is an ϵ^k-perturbed personal equilibrium for every k.

[10] Root nodes—events that are exogenous and not caused by other events in the DAG—are included unconditionally.

[11] It is also instructive to compare expression (6) with the standard chain rule:

$$p(x) \equiv p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_1, x_2) \dots p(x_n \mid x_1, x_2, \dots, x_{n-1}).$$

See Spiegler (2016) for a more detailed discussion.

C Additional Analysis for Baseline Model

C.1 Misperception vs. Reality

In our model, the citizen has incorrect beliefs about the data generating process, believing the data was generated from the *SN*, rather than the *BO* DAG. To emphasize the importance of this assumption, it is useful to contrast the results obtained above with the case in which the true data generating process is given by the *SN*.

Contrary to the model in the main text, suppose that $\Pr(\theta = 1 \mid a) = \gamma$ and $\Pr(y = 1 \mid \theta) = \theta$, i.e., the leader is strong with probability γ if the citizen supports and economic performance is good if and only if the leader is strong. As the *SN* specifies, citizen support is the only cause for leader strength, which in turn causes economic performance.

Suppose that the citizen's utility function is $y + ax$ as before, and the citizen believes in the *SN*, which here means that she has rational expectations. Then, the expected utility of supporting the leader is $\gamma + x$ because given support, the leader will be strong with probability γ, and high economic performance is realized with probability 1. In addition, the citizen obtains the ideological benefits x. By contrast, expected utility of not supporting the leader is 0 because without support, the leader will be weak for sure, and economic performance will be poor. The citizen hence supports if $x + \gamma \geq 0$, i.e., if the ideological benefits are high enough.

This decision rule is very different from the one in which the citizen has incorrect beliefs, as analyzed in the main text. In particular, here, the net expected utility of supporting is *independent* of past behavior and there is always a unique threshold for supporting the leader ($x \geq -\gamma$). We have shown in the main text that when the citizen believes in the *SN* and true data generating process is given by the *BO* DAG, the net expected utility depends on past behavior and there are multiple (personal) equilibria.

C.2 Propaganda Effort

We briefly investigate a leader's incentives to exert "propaganda effort," denoted by $e \in [0, 1]$, to convince the citizen that the *SN* is the true data generating process. We

assume that with probability e, the citizen employs the *SN*. With probability $1-e$, the citizen has rational expectations and hence employs the *BO*. The incumbent's utility function is:

$$U_I = a + \omega_\theta \theta - c(e),$$

where ω_θ is the concern for strength and c is the cost function for effort.

Using the results derived in the main text, and denoting by x^{SN} the equilibrium threshold of support, where $x^{SN} \in \{x_0, x_1\}$ from Proposition 1, the incumbent's optimization problem is:

$$\max_{e\in[0,1]} \left[e\left(1-F(x^{SN})\right) + (1-e)\left(1-F(x^{BO})\right)\right](1+\omega_\theta\gamma) - c(e)$$

The first-order condition is:

$$\left(F(x^{BO}) - F(x^{SN})\right)(1+\omega_\theta\gamma) - c'(e) = 0$$

The result mentioned in the main text then follows from Proposition 2.

D Additional Analysis for Backsliding Model

D.1 Effectiveness of the *SN*

For completeness, we show here that the effect of γ on the effectiveness of the *SN* is again ambiguous.

Suppose first that $\tilde{x}^{SN} = \gamma(\psi q - 1)$. Then, we have:

$$\frac{\partial \text{Eff}^{SN}}{\partial \gamma} = \psi q \left[f(\tilde{x}^{BO}) - f(\tilde{x}^{SN})\right] + f(\tilde{x}^{SN}).$$

This cannot be signed in general.

Now suppose that $\tilde{x}^{SN} = \gamma(\psi q - \frac{1-\gamma}{1-m\gamma})$. Then, we have:

$$\frac{\partial \text{Eff}^{SN}}{\partial \gamma} = \psi q \left[f(\tilde{x}^{BO}) - f(\tilde{x}^{SN}) \right] + f(\tilde{x}^{SN}) \frac{1 - 2\gamma - m\gamma^2}{(1 - m\gamma)^2}.$$

This cannot be signed in general. Note that, because of term $\frac{1-2\gamma-m\gamma^2}{(1-m\gamma)^2}$ can be positive or negative, the effectiveness of the *SN* can be higher or lower when γ increases if even if F is given by a Uniform distribution (so that $f(\tilde{x}^{SN}) = f(\tilde{x}^{SN})$).

D.2 Manipulation when Ideology is Uniformly Distributed

Suppose that $F = \mathcal{U}(-1, 1)$. When the citizen has rational expectation, the first-order condition simplifies to:

$$\frac{\psi\gamma q + 1}{2} (1 + \omega_\theta \gamma + \omega_b \gamma q) = k'(m). \tag{7}$$

When the citizen believes in the *SN*, the first-order condition simplifies to:

$$\left(-\frac{1}{2}(1 - m) \frac{\partial \tilde{x}^{SN}}{\partial m} + \frac{\tilde{x}^{SN} + 1}{2} \right) (1 + \omega_\theta \gamma + \omega_b \gamma q) = k'(m). \tag{8}$$

Suppose first that $\tilde{x}^{SN} = \tilde{x}_0 = \gamma(\psi q - 1)$, which is independent of m. The left-hand of Equation 7 is strictly larger than the left-hand side of Equation 8. Hence, it must be the case that the optimal level of manipulation is lower when the citizen believes in the *SN*.

Now suppose that $\tilde{x}^{SN} = \tilde{x}_1 = \gamma \left(\psi q - \frac{1-\gamma}{1-\gamma m} \right)$, which is decreasing in m. Specifically:

$$\frac{\partial \tilde{x}_1}{\partial m} = -\frac{\gamma^2(1 - \gamma)}{(1 - m\gamma)^2}.$$

Using this expression, the left-hand of Equation 7 is strictly larger than the left-hand side of Equation 8 if:

$$\frac{\psi\gamma q + 1}{2} > \frac{1}{2}(1 - m) \frac{\gamma^2(1 - \gamma)}{(1 - m\gamma)^2} + \frac{\gamma \left(\psi q - \frac{1-\gamma}{1-\gamma m} \right) + 1}{2}$$

Simplifying, this always holds. Hence, it must be the case that the optimal level of manipulation is lower when the citizen believes in the *SN*.