

Media Freedom, Nationalist Protests, and Audience Costs*

David Krosin[†] and Benjamin Shaver[‡]

2024-08-09

Abstract

Why do authoritarian regimes allow media dissent during international crises? Extant literature suggests that media freedom is costly for authoritarian regimes, especially during a crisis, and that regimes may benefit from censorship and propaganda to engender anti-foreign sentiment during international conflict. We explore this puzzle using a model where an autocrat uses her control of the media to persuade citizens to participate in anti-foreign protests that act as a signal of resolve to a foreign enemy. However, these anti-foreign protests can shift to anti-regime protests with some probability that is increasing in the protest's size. We characterize the conditions under which the autocrat balances the benefits and costs of allowing protests and find that the generation of audience costs via nationalist protests is simultaneously a soft and blunt tool that is only used in certain circumstances as the risk it poses to the regime often outweighs the benefits.

*Thanks to Amartya Bose, Christian Godwin, Robert Gulotty, Issac LaGrand, Andrew Little, John Patty, Konstantin Sonin, Hani Warith, and participants at a Summer 2020 Formal Theory Society Junior Scholar Symposium for helpful feedback.

[†]Data Analyst, Thrive Chicago. Krosind@gmail.com

[‡]Ph.D. Student, Irving B. Harris School of Public Policy, The University of Chicago. Blshaver@uchicago.edu

1 Introduction

Why would an autocrat allow some degree of media freedom during an international crisis? Extant literature has established a strong link between media freedom and anti-regime behavior and thus tends to emphasize the costs autocrats face by allowing freer media. For example, previous work suggests that increased media freedom can facilitate citizens learning about poor government performance and collectively organizing (Egorov and Sonin 2009; Enikolopov et al. 2011; King et al. 2013; Little 2016; Lorentzen 2014; Manacorda and Tesei 2020). The literature also emphasizes that while foreign policy crises are high stakes events, media coverage has significant latitude to shape the opinions and attention of citizens (Pan and Kosicki 1994; Soroka 2003). This, combined with the fact that international crises may provide an opening for large-scale mobilization against the regime, suggests that international crises are particularly risky times to reduce media control (Cairns and Plantan 2016; King et al. 2013). Yet, despite work that suggests autocrats will maintain biased media during international crises, we observe examples to the contrary.

In this paper, we explore this theoretical puzzle by drawing on a burgeoning literature that asserts that during instances of crisis bargaining, the presence of nationalist protests in autocratic countries sends a credible signal of resolve to foreign actors by introducing a degree of risk to the autocrat’s survival in office (Li and Chen 2020, Weeks 2012; Weiss 2013, 2014, Weiss and Dafoe 2019). In contrast to much of the literature on crisis-bargaining which asserts that non-democracies struggle to generate audience cost, this growing literature provides evidence that nationalist protests in non-democracies function much like elections democracies to generate risk to the leader’s tenure that conveys their intention not to back down.¹

To explore this puzzle, we formalize a model in which an autocrat may generate audience costs in the form of nationalist, anti-foreign protests to induce a foreign adversary to back down. To this end, the autocrat attempts to persuade citizens to participate in such protests by setting the amount of nationalist bias in the media. This shapes what citizens learn about the autocrat’s valuation of the conflict at hand, potentially altering their individual decisions to participate in the protests.

Across this framework, we develop the intuition that generation of autocratic audience costs via nationalist protests is a simultaneously blunt and soft tool that is constrained in its usage by many factors. Any time an autocrat attempts to generate anti-foreign protests to produce audience costs, she also generate risk to their regime. Namely, nationalist protests can easily become anti-regime protests, which are costly to put down. Additionally, while larger protests may be more effective from the standpoint of generating audience costs, as the size of the nationalist protests increases, so does the risk that the protests will shift from anti-foreign to anti-regime. The latent threat posed by the citizens’ ultimately provides an upper bound on the effectiveness of nationalist protests as audience costs, as the autocrat will generate nationalist protests only if the benefit

of generating one outweighs the potential costs, and will actively dissuade protests when the cost outweighs the benefit. This finding lends itself to the empirical observation that nationalist protests are not always present when autocrats are engaged in international conflict.

In developing our model, we bridge a range of literatures. First, we engage with the large international relations literature that examines how states generate audience costs while crisis bargaining. This discussion stems from Fearon's (1994) work that formalizes the argument that when leaders back down during an international crisis, they pay an audience cost. As a result, leaders are able to generate costly signals by issuing threats and statements of resolve in front of a domestic audience who will punish the leader if they back down. Within the audience cost literature, our work is closest to Slantchev (2006) who models how citizens learn about policy quality—either through politically motivated behavior (the political opposition) or through a noisy and possibly biased signal (the media). Relative to Slantchev (2006), we differentiate ourselves in two ways. First, we characterize how the level of media freedom in the coverage of the conflict is endogeneously determined. Second, we require citizens to opt into consuming the news, which constrains the autocrat as they need to make the media informative enough that citizens tune in. Our paper also builds on work by Baum and Potter (2008, 2015) and Potter and Baum (2010, 2014) who examine how media freedom and media access affect how citizens learn about foreign policy in democracy.

Our framework also builds on a growing literature that points to nationalist protests as a mechanism by which authoritarian governments are able to generate audience costs (Ding 2018; Ma 2020; Weeks 2008; Weiss 2013; Weiss 2014; Weiss and Dafoe 2019). Relative to this work, we distinguish ourselves along multiple dimensions. First, we allow the size and severity of nationalist protests to be endogenous to the regime's control of the media and their commitment to reporting the truth about the autocrat's type with a sufficiently high probability. Extant literature considers the size of the protests to be exogenously determined by preexisting nationalist sentiment that can be encouraged or repressed (Weiss 2013, 2014). A recent exception is Ding (2018) who builds a crisis bargaining model characterizing when a regime has incentives to promote costly protests where the foreign audience is uncertain about the identity of the protesters. Our model differs from Ding's in several respects. First, we endogenize domestic media censorship and characterize a mechanism by which autocrats can promote nationalist sentiment in a way that is credible to foreign audiences. Second, our model incorporates the risk that protests have for regime and leader survival. The credibility of the autocrat's signal depends on her sensitivity to protests and the possibility that they turn anti-regime. Ultimately, our model predicts that autocrats will not allow any media freedom if the risk to the regime is too great. Third, our model implies that autocrats can generate protests and credibly signal resolve by reducing anti-foreign media bias and increasing media freedom. Extant literature on anti-foreign protests argue that they are fueled by nationalist propaganda (e.g. Mattingly and Yao 2020). By contrast,

in our model, sufficiently biased media is not informative and unable to persuade citizens to protest.

Finally, this paper is also closely related to the growing theoretical literature that applies models of Bayesian persuasion to examine how autocrats utilize information disclosure in weakly institutionalized environments (Boleslavsky, Shadmehr, and Sonin 2021; Luo and Rozenas 2018; Shadmehr and Bernhardt 2015).² Our work is similar to Gehlbach and Sonin (2014), who develop a model in which a government selects the optimal level of media bias to persuade citizens to take a certain action and to maximize profits from the media firm. Similar to their work, we develop a model in which Bayesian citizens decide whether to consume the news. Then, those citizens who do watch the news use information reported when making a costly decision whose outcome depends on the state of the world. We distinguish ourselves relative to this paper by unifying our model with the expansive international relations literature that attempts to understand audience costs, particularly whether autocratic countries have the ability to use them, and the relationship of audience costs to nationalist protest. We also differentiate ourselves by considering a model where the autocrat faces a trade off when generating protests. Instead of maximizing advertising revenue or the probability citizens take the autocrat's preferred action, the autocrat in our model must balance the benefit gained from increasing the probability the enemy will back down with the risk stemming from large protests.

The paper proceeds as follows. First, we motivate the paper by discussing the three components we formalize in a model. Second, we formalize our model and derive comparative statics. Finally, we conclude with a summary of our findings and suggestions for the direction of future work.

2 Motivation

In this paper we incorporate three conceptual elements into a formal framework: audience costs, nationalist protests, and media bias. We briefly detail the motivations for these elements in this section.

2.1 Audience Costs

During international crises, leaders often struggle to convey their intentions or level of resolve through costless statements. One way that leaders work around this is by generating audience costs (Fearon 1994). By publicly issuing a threat or statement of resolve in front of their domestic audience, leaders generate audience costs by increasing the cost of backing down. Because domestic audiences will punish a leader if they are perceived to be bad at foreign policy leaders can increase the credibility of their threats and statements of resolve by generating audience costs.

The ability to generate audience costs hinges on two factors. First, citizens must have the ability to

punish their leader if they back down, as the threat of this is what changes the leader’s statement from cheap talk to a costly signal. Second, the citizens’ punishment mechanism must be observable to the actor the state is crisis bargaining with. In the absence of one or both of these factors, the leader is not able to generate audience costs. At a first pass, this implies that democratically elected leaders will be more capable of generating audience costs than authoritarian leaders, as democratic elections satisfy both components (Fearon 1994). Much of the audience cost literature operates under this assumption, resulting in a dearth of work on audience costs in countries led by autocrats (Eyerman and Hart 1996; Gelpi and Griesdorf 2001; Schultz 2001).

However, recent work has begun to push back on this intuition by highlighting ways in which autocrats are susceptible to punishment. For example, Bueno de Mesquita and et. al. (2003) argue that autocratic leaders serve at the whim of a selectorate or winning coalition which is comprised of elites in the military, a political party, or the bureaucracy. This means that autocrats are not monolithic—they still may pay costs if they show themselves to be ineffective at conducting foreign policy. Similarly, Weeks (2012) suggests that in some autocracies institutional structures open leaders up to punishment for poor foreign policy, with heterogeneity across multiple dimensions explaining the observed variation in the ability in autocrats to successfully generate audience costs. In the model below, the first concept we formalize is the ability of an autocratic leader to generate audience costs by generating a pathway through which they can be punished for conducting poor foreign policy.

2.2 Nationalist Protests

However, it is not enough for an autocrat to face risk to render their statements costly. To generate audience costs that can result in a costly signal, the risk the autocrat faces must be observable to the audience they are engaged in crisis bargaining with. One way that autocrats are able to do this is by giving the green light to nationalist, anti-foreign street protests, authoritarian leaders can credibly invoke the threat of public opinion while engaged in crisis bargaining (Weiss 2013, 2014).³ This is the second component of the model we develop below.

Nationalist protests introduce risk to the autocrat’s regime along two pathways. The first is simply through the increased risk any type of large scale citizen mobilization poses to an autocrat’s regime. For many reasons, allowing citizens to mobilize introduces risk that actors will work to remove the autocrat, even if the citizens ostensibly are mobilizing in opposition against a foreign actor. First, anti-foreign protests can act as tipping points or information cascades. As citizens learn that protesting is safe or encouraged, they will join the streets. Once a tipping point is reached, the protests can begin to swell rapidly (Kuran

1991). Second, nationalist protests can lower the cost of collective action, allowing previously passive citizens to join protests at a lower cost than they would face otherwise. Third, protests can lead to elite schisms within the autocrat's selectorate as protests may expose government weakness (O'Donnell and Schmitter 1986; Przeworski 1991). Finally, nationalist protests typically occur in large cities, where dissent can be particularly harmful to autocratic survival due to the lower costs for collective action (Wallace 2013, 2014).⁴

The second pathway is that nationalist protests also increase the cost of backing down for the autocrat. Much like citizens in countries with democratic elections who punish their leaders by voting them out, anti-foreign protests can quickly become anti-regime protests if the autocrat backs down after issuing a threat and is perceived as having made a mistake. Even if the shift from anti-foreign to anti-regime doesn't result in the autocrat being removed from office, it will be costly for them to put down the protests. Consider, the case of Chinese General Secretary Hu Yaobang (Weiss 2014, pp. 99-101). Nationalist protests turned anti-regime in 1986 following what was considered a weak response by the Chinese government to anti-Japanese protests in 1985. In response, party elites removed Hu from office.⁵ As the size of protests increases, the likelihood they turn into anti-regime protests increases. Similarly, if the autocrat backs down after issuing a threat, they risk the anti-foreign protests becoming anti-regime.

The 1992 removal of President Ayaz Mutallibov in Azerbaijan provides a further illustrative case for how nationalist protests can turn anti-regime and lead to leader removal. Over the course of the previous four years, Armenia and Azerbaijan had been engaged in territorial conflict where Azerbaijan was attempting to stymie the secessionist movement of the Armenian majority territory Nagorno-Karabakh and Armenia was supporting their co-ethnics in the conflict. With the fall of the Soviet Union the conflict escalated and Armenia launched an offensive that culminated in the capture and subsequent massacre of the Azerbaijani town of Khojaly resulting in hundreds of civilian deaths. Nationalist, anti-Armenian demonstrations raged in the Azerbaijani capital Baku calling for an aggressive response and widening of the conflict.⁶ When Mutallibov was seen to take a conciliatory position, claiming that death tolls following the Khojaly massacre were inflated by opposition parties, protesters began calling for his resignation to which he obliged within the week (Ipek 2009).⁷ In this case, protests increased scrutiny over the leader's actions in response to the massacre which eventually led to calls for his removal even as he was preparing for a counter-offensive, implying that even if a leader obliges in escalating the conflict, it may not be enough to subdue popular outrage.

2.3 Nationalist Media Bias and Mobilization

The extant empirical literature suggests that media with nationalist bias can produce nationalist animosity among citizens who consume it (DellaVigna et. al. 2014, Mattingly and Yao 2020). Biased media also can induce action in those who consume it (Barber and Miller 2019). For example, Enikolopov and coauthors (2011) find that access to independent news resulted in support for opposition parties in the 1999 parliamentary elections in Russia.⁸ This outlines the final component of the formal model: the use of nationalist bias in the media to mobilize citizens. As is often assumed in the formal literature on media bias, we assume that state censorship can be used to create supply-driven media bias that misreports or covers up information that is detrimental to the state. In the model below, we specifically consider nationalist bias.

3 Model

This section formalizes a theory of how autocrats utilize control of the media to gain leverage while crisis bargaining. To do this, we develop a crisis bargaining model using a Bayesian persuasion framework.⁹ Consider a strategic interaction with three players: an autocrat (she), a foreign enemy (he), and a group of citizens in the autocrat's country. The autocrat and the foreign enemy are engaged in a crisis bargaining game where the probability the foreign enemy will back down is increasing in the size of nationalist protests he observes. To induce protests, the autocrat attempts to shape the citizens' beliefs about her valuation of the conflict to prompt their mobilization in opposition to the foreign enemy. Rather than simply choosing the message that the media reports directly, however, the autocrat uses her control of the media to shape the flow of information about the nature of the conflict to the citizen with the actual message determined probabilistically.¹⁰ Consistent with the principle outlined in Section 2.2, any nationalist protests that occur evolve into anti-regime protests with some probability that is increasing in the size of nationalist protests. Intuitively, the autocrat must weigh the trade off between sending a costly signal of resolve to the foreign adversary with the risk of generating protests that may turn against the regime.

3.1 Structure

Consider a model with three players: an autocrat (A , she), a non-strategic foreign enemy (F , he), and a continuum of citizens with mass one (indexed by $i \in [0,1]$). The autocrat is one of two types $s \in \{0,1\}$ where $s = 1$ refers to an autocrat that has a high valuation for the issue in conflict with the foreign enemy, while $s = 0$ denotes an autocrat with a low valuation (Fearon, 1997).¹¹ At the start of the game, the citizens and the autocrat share a common prior that $s = 1$ with probability $\phi \in (0,1)$.¹²

In the first stage of the game the autocrat publicly commits to a media policy $\beta(s) \in [0,1] \times [0,1]$ where $\beta(s)$ is the probability that the media outlet reports a signal of $\hat{s} = 1$ when the state of the world is $s \in \{0,1\}$. Here the autocrat's selection of $\beta(s)$ refers to the level of anti-foreign or nationalist bias in the media that reports about the conflict and the autocrat's valuation. A large literature indicates that individuals are able to detect and filter media bias; therefore, we assume the autocrat's selection of $\beta(s)$ is observable to the citizens.¹³ Once the autocrat selects the level of bias, each citizen individually chooses whether to *watch the news* or *not watch the news*, $c_i \in \{0,1\}$, where watching the news requires paying a cost $\lambda \in [0,1]$.

Next, nature privately reveals the state of the world which is mapped into a signal via the media according to the autocrat's selection of $\beta(s)$. The signal is observed by each citizen who chose to watch the media. Then all citizens individually choose whether to participate in anti-foreign protests or not, $a_i \in \{0,1\}$, where $z = \int_0^1 a_i d_i$ is the total size of protest. The enemy observes the size of the protests and decides whether to *back down* or *escalate to war*, $a_F \in \{0,1\}$, where the probability he will back is $\psi(z)$, which is a weakly increasing, concave, twice continuously differentiable function in z with $\psi(0) = 0$, $\psi(1) = 1$.¹⁴

If the enemy escalates then the autocrat chooses whether *back down* or *escalate to war*, $a_A \in \{0,1\}$. If both the foreign enemy and autocrat escalate then the autocrat wins the ensuing military conflict with probability $p \in (0,1)$. In the final stage of the game, irrespective of whether military conflict occurs, the nationalist protests of size z turns into anti-regime protests with probability $\theta(z)$, where $\theta(z)$ is a weakly increasing, convex, twice continuously differentiable function of z with $\theta(0) = 0$ and $\theta(1) = 1$.¹⁵

To close, the players' receive the following payoffs. If the foreign enemy backs down or the autocrat wins the conflict with the foreign enemy, the autocrat receives a payoff of μ_H (when $s = 1$) or μ_L (when $s = 0$), where $\mu_H > \mu_L > 0$. If the autocrat backs down or loses the conflict with the foreign enemy, she receives a payoff of 0. Regardless, she pays a cost of $k > 0$ if she engages in military conflict. Additionally, if the anti-foreign protests become anti-regime at the end of the game, the autocrat pays a cost of $\pi > 0$. Each citizen i receives $q_i \sim [\bar{q}, 1]$ for not participating in the protests when $s = 0$, where $\bar{q} > \phi$.¹⁶ Citizen i also receive $1 - q_i$ for participating in nationalist protests when $s = 1$. Following Gehlbach and Sonin (2014) and Feddersen and Pesendorfer (1998), the parameter q_i measures citizen i 's aversion to mobilization.

Summarizing, the game proceeds as follows:

1. The autocrat publicly selects a media policy $\beta(s) \in [0,1] \times [0,1]$.
2. Each citizen decides whether to watch the news, $c_i \in \{0,1\}$.
3. The state of the world is realized $s \in \{0,1\}$ and is mapped into a signal $\hat{s} \in \{0,1\}$ according to the autocrat's selection of media policy. The citizens who watch the news observe the signal realization, and all citizens decide whether to participate in an anti-foreign protest.

4. The foreign enemy observes any protests and decides whether to back down or escalate, $a_F \in \{0,1\}$, where the probability he backs down is given by $\psi(z)$, which is increasing in the size of protest.
5. If the foreign enemy chooses not to back down, the autocrat chooses whether to back down or escalate the conflict to war, $a_A \in \{0,1\}$.
6. Any anti-foreign protests become anti-regime protests with probability $\theta(z)$, which is increasing in the size of protests generated.

3.2 Utility Functions

Before solving the model, we explicitly define the autocrat's and citizens' utility functions. The the autocrat's utility function is

$$\begin{aligned} \mathbb{E}[U_A] = & \mathbb{1}_{a_F=0} \left[\phi[\mu_H - \theta(z)\pi] + (1 - \phi)[\mu_L - \theta(z)\pi] \right] \\ & + \mathbb{1}_{a_F=1} \mathbb{1}_{a_A=1} \left[\phi[p\mu_H - k - \theta(z)\pi] + (1 - \phi)[p\mu_L - k - \theta(z)\pi] \right] \\ & - \mathbb{1}_{a_F=1} \mathbb{1}_{a_A=0} \left[\theta(z)\pi \right] \end{aligned}$$

where the indicator variables indicate whether the autocrat or foreign enemy back down or not.

Turning to citizens, if citizen i consumes the media, their expected utility is

$$\mathbb{E}[U_i(c_i = 1)] = \mathbb{1}_{a_i=s=0}[q_i - \lambda] + \mathbb{1}_{a_i=s=1}[(1 - q_i) - \lambda],$$

and if they do not consume the news, their expected utility is

$$\mathbb{E}[U_i(c_i = 0)] = (1 - \phi)q_i.$$

3.3 Autocrat's Decision to Back Down

We solve for a perfect Bayesian equilibrium by backward induction. To do so, we start in the final stage of the game where the autocrat decides whether to back down if the foreign enemy chooses to escalate. If the autocrat has a low valuation and backs down after the foreign enemy escalates her expected utility is

$-\theta(z)\pi$. And if the autocrat has a low valuation and escalates after the foreign enemy escalates her expected utility is $p(\mu_L - k - \theta(z)\pi) + (1-p)(-k - \theta(z)\pi)$. Comparing the two, the autocrat will escalate to military conflict when

$$\mu_L > \frac{k}{p}.$$

The expected utilities for backing down and escalating when the autocrat is the high type can be found by substituting μ_H for μ_L . Therefore, an autocrat with a high valuation will escalate when $\mu_H > \frac{k}{p}$. This leads to the first assumption in our analysis.

Assumption 1 $\mu_H > \frac{k}{p} > \mu_L$.

Assumption 1 implies that an autocrat with a high valuation will engage in military conflict with the enemy, whereas an autocrat with a low valuation will back down in the final stage. Assumption 1 captures one of the fundamental components of crisis bargaining and audience costs. Since military action is costly and not guaranteed to be successful, the only an autocrat who values the issue highly will engage in it. But, if the autocrat can convince the foreign enemy they value the issue highly, even if they don't, the enemy may decide to back down. This uncertainty is what forces the autocrat to utilize a costly signal.

3.4 Citizen's Decision Protest

Having established when an autocrat will prefer to back down in the final stage of the game, we turn to the citizens' decision whether to protest. Citizen i will only protest if $\phi(1 - q_i) > (1 - \phi)q_i$ or $\phi > q_i$. Because we assume that $\bar{q} > \phi$, no citizens protest in the absence of persuasion. This means that none of the citizens who choose not to watch the news will protest.

Let us assume that $P(s = 1|\hat{s} = 1) = \beta(1) = 1$; that is, the autocrat will always choose a media policy that accurately reports the state of the world when $s = 1$.¹⁷ Then the posterior belief of citizen i who watches the news can be summarized,

$$P(s = 1|\hat{s} = 1) = \frac{\phi}{\phi + (1 - \phi)\beta(0)}$$

$$P(s = 1|\hat{s} = 0) = 0$$

We see that the posterior belief that $s = 1$ conditional on $\hat{s} = 1$ is decreasing in $\beta(0)$. That is, the more

the autocrat biases the media, the less likely citizens are to believe what the media reports when it reports $\hat{s} = 1$. Therefore, as bias increases, citizen i learns less from consuming the news (Gehlbach and Sonin 2014).

When citizen i chooses whether to watch the news, they compare the expected utility of consuming the news relative to the expected utility of taking $a_i = 0$, which they will do in the absence of more information. The expected utility of consuming the news is

$$(1 - \phi)(1 - \beta(0))q_i + [\phi + (1 - \phi)\beta(0)](1 - q_i) \left[\frac{\phi}{\phi + (1 - \phi)\beta(0)} \right] - \lambda$$

which is equivalent to $(1 - \phi)(1 - \beta(0))q_i + (1 - q_i)\phi - \lambda$. This can be compared to the expected utility from consuming no media, $(1 - \phi)q_i$. Therefore, the marginal utility of consuming the media for i is

$$(1 - q_i)\phi - \beta(0)(1 - \phi)q_i - \lambda.$$

This means that only citizens for which

$$(1 - q_i)\phi - \beta(0)(1 - \phi)q_i > \lambda$$

will watch the news. This can be rearranged in terms of q_i

$$\frac{\phi - \lambda}{\phi + \beta(0)(1 - \phi)} > q_i$$

This means that all citizens with a value of q_i such that $\frac{\phi - \lambda}{\phi + \beta(0)(1 - \phi)} > q_i$ will choose to watch the news.

Define $\frac{\phi - \lambda}{\phi + \beta(0)(1 - \phi)}$ as q^* . Because q_i is uniformly distributed on $[\bar{q}, 1]$, the expected mass of citizens who watch the news is $M = \max\{\frac{q^* - \bar{q}}{1 - \bar{q}}, 0\}$. Sequential rationality implies that no citizens watch the news who will not protest conditional on observing $\hat{s} = 1$. This means that the mass of citizens who watch the news is equivalent to the mass of citizens who protest when they observe $\hat{s} = 1$; therefore we may also define the level of protests conditional on $\hat{s} = 1$ as $z = M$. This means that when the expected mass of citizens who will protest is positive, it is increasing in ϕ , and decreasing in $\beta(0)$ and λ .

Additionally, since z is decreasing in $\beta(0)$, z is largest when $\beta(0)$ is 0. To ensure that it is possible for the autocrat to generate positive level of protests, Assumption 2 ensures that when $\beta(0) = 0$, $z > 0$.

Assumption 2 $\phi(1 - \bar{q}) > \lambda$.

If this assumption does not hold, which happens when λ is sufficiently large or ϕ is sufficiently small, the autocrat is unable to ever generate positive levels of protests. This assumption suggests that the autocrat's ability to generate protests may be constrained by factors such as the cost citizens pay to consume the news or the citizens' beliefs about the autocrat's valuation for the issue in conflict. We will explore this in more depth below.

3.5 Autocrat's Optimal Media Policy

Having derived the expected mass of citizens who will protest conditional upon the media reporting $\hat{s} = 1$, we continue with backwards induction and now solve for the autocrat's optimal media policy. The autocrat's optimal strategy is derived by solving the following

$$\begin{aligned} \max_{\beta(0) \in [0,1]} & \underbrace{(\phi + (1 - \phi)\beta(0))}_{P(\hat{s}=1)} \psi \left(\frac{\frac{\phi - \lambda}{\phi + (1 - \phi)\beta(0)} - \bar{q}}{1 - \bar{q}} \right) \left[\phi(\mu_H - \theta \left(\frac{\phi - \lambda}{\phi + (1 - \phi)\beta(0)} - \bar{q} \right) \pi) \right. \\ & \left. + (1 - \phi)(\mu_L - \theta \left(\frac{\phi - \lambda}{\phi + (1 - \phi)\beta(0)} - \bar{q} \right) \pi) \right] + \underbrace{(\phi + (1 - \phi)\beta(0))}_{P(\hat{s}=1)} (1 - \psi \left(\frac{\phi - \lambda}{\phi + (1 - \phi)\beta(0)} - \bar{q} \right)) \\ & \left[\phi(p\mu_H - k) - \theta(z)\pi \right] + \underbrace{(1 - \phi)(1 - \beta(0))}_{P(\hat{s}=0)} \left[\phi(p\mu_H - k) \right]. \end{aligned}$$

Which simplifies to

$$\begin{aligned} \max_{\beta(0) \in [0,1]} & (\phi + (1 - \phi)\beta(0)) \left[\psi(z)(\phi\mu_H + (1 - \phi)\mu_L) + (1 - \psi(z))\phi(p\mu_H - k) - \theta(z)\pi \right] \\ & + (1 - \phi)(1 - \beta(0)) \left[\phi(p\mu_H - k) \right]. \quad (1) \end{aligned}$$

The first term in the equation 1 is the probability that $\hat{s} = 1$ multiplied by the payoff of generating anti-foreign protests. When protests are successful in inducing the enemy to back down, then the autocrat receives utility corresponding to their expected valuation of winning the crisis, $\phi\mu_H + (1 - \phi)\mu_L$. But protests may also be unsuccessful in inducing the enemy to back down (with probability $1 - \psi(z)$), resulting in the autocrat receiving his expected payoff when the enemy escalates the conflict, $\phi(p\mu_H - k)$. Regardless of whether the enemy back's down, the autocrat will face anti-regime protests with some probability ($\theta(z)$) and pay a cost if anti-regime protests are realized (π). The second component is the probability the signal $\hat{s} = 0$ multiplied by the expected payoff when the enemy escalates the conflict. In this component, citizens observe

that the autocrat is low resolve, no protests occur, and the enemy escalates with certainty.

The answer to the maximization problem is provided in Proposition 1.

Proposition 1 *When there exists some $z \in \left(0, \frac{\phi-\lambda}{1-\bar{q}}\right)$ such that $\frac{\phi(1-\bar{q})}{\phi-\lambda} < \frac{\psi'(z)\Sigma - \theta'(z)\pi}{\psi(z)\Sigma - \theta(z)\pi} < \frac{1-\bar{q}}{\phi-\lambda}$, the autocrat's equilibrium strategy is to generate expected protests of size z^* by setting bias to*

$$\beta(0) = \frac{(\phi - \lambda) \left[\psi'(z^*)\Sigma - \theta'(z^*)\pi \right] - \phi(1 - \bar{q}) \left[\psi(z^*)\Sigma - \theta(z^*)\pi \right]}{(1 - \phi)(1 - \bar{q}) \left[\psi(z^*)\Sigma - \theta(z^*)\pi \right]},$$

where $\Sigma = \phi\mu_H + (1 - \phi)\mu_L - \phi(p\mu_H - k)$

Otherwise, the autocrat's optimal policy is to set $\beta(0) = 0$ when

$$\psi\left(\frac{\phi-\lambda}{1-\bar{q}}\right)\Sigma > \theta\left(\frac{\phi-\lambda}{1-\bar{q}}\right)\pi,$$

to set $\beta(0) = 1$ when

$$\psi\left(\frac{\phi-\lambda}{1-\bar{q}}\right)\Sigma < \theta\left(\frac{\phi-\lambda}{1-\bar{q}}\right)\pi,$$

and to choose either when

$$\psi\left(\frac{\phi-\lambda}{1-\bar{q}}\right)\Sigma = \theta\left(\frac{\phi-\lambda}{1-\bar{q}}\right)\pi.$$

Proof. All formal proofs are in the appendix. ■

Proposition 1 says that the autocrat will not always try generate protests. When there is not some $z \in \left(0, \frac{\phi-\lambda}{1-\bar{q}}\right)$ such that $\frac{\phi(1-\bar{q})}{\phi-\lambda} < \frac{\psi'(z)\Sigma - \theta'(z)\pi}{\psi(z)\Sigma - \theta(z)\pi} < \frac{1-\bar{q}}{\phi-\lambda}$, and when $\psi\left(\frac{\phi-\lambda}{1-\bar{q}}\right)\Sigma < \theta\left(\frac{\phi-\lambda}{1-\bar{q}}\right)\pi$ the autocrat's expected utility is higher from generating no protests than it is from trying to generate protests of any size. This may happen for a number of reasons. For example, Σ may be small (or negative), which occurs when the cost of conflict is high or when the the autocrat's valuation for the issue is small. Or, π may be large, which happens when the cost of anti-regime protests is very high. When this happens the autocrat will set $\beta(0) = 1$ which means no citizens protest because the media is so uninformative that no citizens watch the news.

Proposition 1 also says that when there is an interior solution, the autocrat chooses $\beta(0)$ to balance the benefits of inducing protests (increasing the probability the foreign enemy backs down) with the costs (increasing the probability that an anti-foreign protests become anti-regime protests). Corollary 1 states this more explicitly by demonstrating that when $\beta(0) \in (0,1)$, $\beta(0)$ is chosen so that the marginal benefit of the protests is equal to the marginal cost of the protests.

Corollary 1 *When the optimal $\beta(0)$ is interior, $\beta(0)$ is chosen to solve so that the following ratio is satisfied*

$$\frac{\Sigma}{\pi} = \frac{P(\hat{s} = 1)z'\theta'(z) + (1 - \phi)\theta(z)}{P(\hat{s} = 1)z'\psi'(z) + (1 - \phi)\psi(z)}.$$

Having derived the optimal level of bias, we now derive comparative statics on the parameters in the model. Assume that the autocrat chooses a level of bias that is interior. Proposition 2 explores how that level of bias changes in the other parameters.

Proposition 2 *Assume that the autocrat chooses a level of bias that is interior and therefore generates an expected level of bias $z^* \in \left(0, \frac{\phi - \lambda - \bar{q}}{1 - \bar{q}}\right)$. Then*

- i. $\beta(0)^*$ is decreasing in the cost to consume the media, λ .*
- ii. $\beta(0)^*$ is decreasing in the lower bound of the benefit to not protesting, \bar{q} .*
- iii. $\beta(0)^*$ is decreasing in the benefit of generating a protest, Σ .*
- iv. $\beta(0)^*$ is increasing in the cost of an anti-regime protest, π .*

Proposition 2 says that when there is some level of protests z^* that the autocrat wants to induce, if the cost to consume the media increase, the autocrat needs to decrease the level of bias to compensate. This is because the number of citizens who watch the news is decreasing in bias and in the cost to consume the media. Therefore, if the cost increases, the autocrat will need to decrease bias to compensate so that z^* protests occur. As a result of decreasing $\beta^*(0)$, $\hat{s} = 1$ less frequently, which means protests will happen less frequently and the foreign enemy will be more likely to escalate the conflict.

Then, Proposition 2 says that the level of bias is decreasing in \bar{q} . A higher level of \bar{q} implies that a greater number of citizens weigh the payoff of not protesting when it is favorable to do so more than the payoff of protesting when it is favorable to do so. This could be when the opportunity cost of protesting is relatively high and even the citizens who are most likely to protest relative to other citizens have high reservations about mobilizing. For example, if a leader has historically punished public demonstrations, citizens might

be reluctant to mobilize even if the autocrat wants them to. When this occurs, the autocrat must open the media up even more to persuade citizens to protest.

Finally, Proposition 2 says that when the benefit of generating protest increases, Σ , the autocrat will decrease the level of bias. This is because bias reduces the number of citizens who watch the news, which in turn reduces the number of protesters. Therefore, when the benefit of generating protests is large, increasing the number of protesters is beneficial. To do so, the autocrat will reduce bias, which means more protesters will protest, but protests will occur less frequently as $P(\hat{s} = 1)$ decreases. This dynamic can be seen in Figure 1. The leftmost panel shows the $\beta(0)^*$ that solves the ratio in Corollary 1. In the center panel, Σ increases, which shifts the intersection higher, which means a smaller $\beta(0)^*$ solves the ratio.

In a similar dynamic, as the cost of generating protests increases, π , the level of bias increases. The potential costs associated with protests turning anti-regime reduces the autocrat's incentive to produce large protests. To ensure that smaller protests happen, the autocrat increases bias, which decreases the number of citizens watching the news, which then decreases the number of citizens who protest. This can be seen in the rightmost panel of Figure 1.

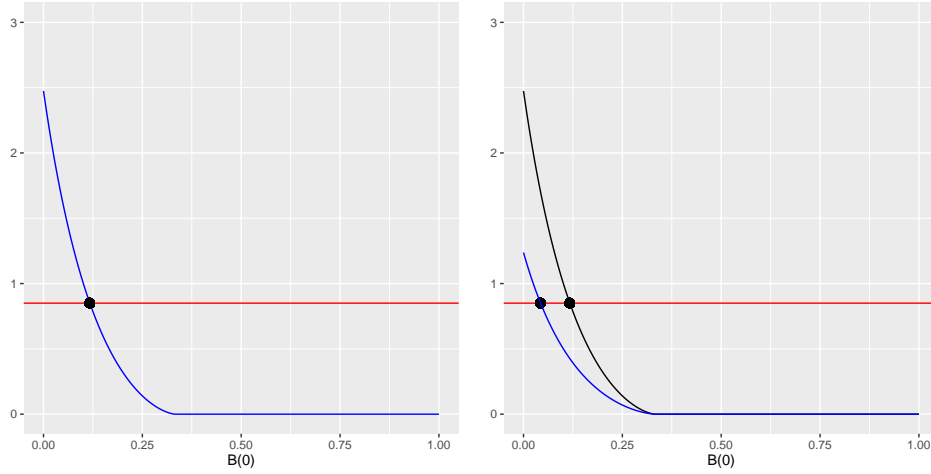
Corollary 2 leverages Proposition 2 to explore the relationship between bias and the parameters in Σ .

Corollary 2 *Assume that the autocrat chooses a level of bias that is interior and therefore generates an expected level of bias $z^* \in \left(0, \frac{\frac{\phi-\lambda}{1-\bar{q}}}{1-\bar{q}}\right)$. Then*

- i. $\beta(0)^*$ is decreasing in the payoff the high type gets from winning the conflict with the foreign enemy, μ_H .*
- ii. $\beta(0)^*$ is decreasing in the payoff the low type gets from winning the conflict with the foreign enemy, μ_L .*
- iii. $\beta(0)^*$ is increasing in the probability the autocrat wins a military conflict with the foreign enemy, p .*
- iv. $\beta(0)^*$ is decreasing in the cost the the autocrat pays to engage in military conflict with the foreign enemy, k .*

Corollary 2 builds on the comparative static for Σ from Proposition 2. Since an increase in Σ leads to less bias in equilibrium, changes to parameters that increase Σ also lead to less bias. When the autocrat's valuation for winning the crisis increases, she decreases bias to generate larger protests. The same is true as the cost for engaging in military conflict with the foreign enemy increases. But as the probability that the autocrat can win a military conflict with the foreign enemy increase, the relative benefit from generating protests decreases, which means the autocrat will increase bias. This is because the autocrat will be more willing to win the conflict outright rather than take on the additional risk posed by anti-regime protests.

Figure 1: Proposition 3



Note: In the leftmost panel $\phi = 0.5, \lambda = 0.1, \bar{q} = 0.6, \mu_H = 2, \mu_L = 0.5, p = 0.6, k = 0.4, \pi = 1, \lambda = 0.1, \Sigma = 0.8, \psi(z) = \sqrt{z}$, and $\theta(z) = z^2$. In the rightmost panel $\psi(z) = 2\sqrt{z}$.

Next, we will examine the effect of $\psi(z)$ and $\theta(z)$ on the equilibrium level of bias.

Lemma 1 *The expected marginal benefit of protests, $\Sigma[P(\hat{s} = 1)z'\psi'(z) + (1 - \phi)\psi(z)]$, is increasing in $\beta(0)$. The expected marginal cost of protests, $\pi[P(\hat{s} = 1)z'\theta'(z) + (1 - \phi)\theta(z)]$, is decreasing in $\beta(0)$*

The intuition behind Lemma 1 is that, given $z' < 0$, the rate of change in the probability that the enemy backs down with respect to protests size, $\psi'(z)$ is decreasing and tending towards 0 while $\psi(z)$ is increasing. But $\theta'(z)$ is larger and increasing at a faster rate than $\theta(z)$.

Proposition 3 *Assume that the autocrat chooses a level of bias that is interior and therefore generates an expected level of bias $z^* \in \left(0, \frac{\phi - \lambda - \bar{q}}{1 - \bar{q}}\right)$. Then*

- i. $\beta(0)^*$ is decreasing as the function for the marginal probability that the enemy backs down, $P(\hat{s} = 1)z'\psi'(z) + (1 - \phi)\psi(z)$, increases for $\beta(0) \in (0, 1)$.*
- ii. $\beta(0)^*$ is increasing as the function for the marginal probability that protests turn anti-regime, $P(\hat{s} = 1)z'\theta'(z) + (1 - \phi)\theta(z)$, increases for $\beta(0) \in (0, 1)$*

Proposition 3 says that when the autocrat is less likely to face anti-regime protests for all levels of protest, then she can afford to allow larger protests and allow more media freedom. Likewise, if the autocrat can induce the enemy to back down at a higher probability for all levels of protests, then she has a greater expected payoff for inducing protests. This last point, in particular, captures an important logic of the audience cost literature. When facing a foreign adversary that can itself cannot generate high audience costs, then the home government is in better position to force him to back down (Fearon 1994).

But changes to the shape of probability functions $\phi(z)$ and $\theta(z)$, such as via changes to their second derivatives, have less obvious effects on equilibrium media freedom. Consider the first derivative of the autocrat’s utility with respect to $\beta(0)$:

$$U'_A = P(\hat{s} = 1)z'(\psi'(z)\Sigma - \theta'(z)\pi) + (1 - \phi)(\psi(z)\Sigma - \theta(z)\pi)$$

Increasing $\beta(0)$ causes a negative change in the first component and a positive change in the second component. These opposing effects are due to the countervailing effect of $\beta(0)$ on the probability $\hat{s} = 1$ and the likelihood of the enemy backing down net the probability of being overthrown. So, a change in the second derivative of either probability function has an ambiguous effect on $\beta(0)^*$ depending on which component effect dominates. For example, if the a change in $\psi(z)$ makes it relatively steep at $\beta(0)^*$ then equilibrium bias will decrease. While a relatively flatter slope will increase bias. Likewise, a relatively steeper $\theta(z)$ will increase bias.

3.6 Discussion

The analysis of this model highlights several points about the ability of an autocrat to generate audience costs via nationalist protest. Namely, the autocrat’s ability to use nationalist protests to generate audience costs is constrained by four factors. First, generating audience costs through anti-foreign protests is a “blunt tool.” If the autocrat drums up anti-foreign protests, they also increase their risk from ant-regime protests. Therefore, the autocrat is constrained by the risk of anti-foreign protests turning anti-regime. This mirrors the trade-offs highlighted in other works on media control in autocracies where a leader must weigh the risk of allowing media freedom with some benefit (Qin et. al. 2017, 2019).

Second, the foreign enemy can require audience costs that are sufficiently high that no selection of media bias can generate protests of a large enough size to induce him to back down. This suggests the manipulation of the domestic media is a “soft tool” during an international crisis. The scholarly literature has identified several ways in which autocratic regimes can generate audience costs, such as through institutional constraints (Weeks 2012) or nationalist protests (Weiss 2013, 2014), when it had previously been theorized that democratic states had an advantage in international crises. But given these methods, the literature has had less to say about why autocrat’s still do not always generate audience costs (Hyde and Saunders 2020). Our model provides some novel insights as to why autocrats cannot generate audience costs at will during international crises, namely that their methods mobilization have limits and may not be sufficient to induce adversaries to back down. This essentially occurs when an interior solution is not possible and the maximum

allowable protests are insufficient to induce the enemy to back down with a high enough probability relative to the risk posed by generating large protests.

Third, the autocrat can only persuade those citizens who watch the news to protest. Therefore, when the cost to consume the news is large, there is a ceiling on how many citizens the autocrat can turn out to protest. So even if the autocrat wanted to turn out large protests, she may not be able to if few citizens consume the news. Furthermore, if media consumption costs are sufficiently high, then no interior solution to the autocrat's optimal media policy problem will exist, leading the autocrat to choose between one of the two corner solutions. But since the maximum size of protests (which occurs under full media freedom) is decreasing in the cost of consumption, the autocrat is increasingly likely to completely bias the media as consumption costs increase. In total, media freedom at first increases with λ to compensate for higher cost of consumption, but if it increases too high then the autocrat will switch to complete media bias.

Fourth, an autocrat may not allow media freedom and generate nationalist protests if she is likely to win a conflict if it is escalated or if she has low valuation for the issue at hand. Taking on the risk of being overthrown is only worth it if it is better than the alternative of the enemy backing down with certainty. But the autocrat faces opposite incentives when the costs of war are increased. This suggests that as the costs of conquest continue to increase as they have in the post WWII period, the use of nationalist protests will become more prevalent except in regimes that are most sensitive to citizen protests.

4 Conclusion

In this paper, we develop a model where an autocrat commits to a media policy to generate nationalist, anti-foreign protests in an effort to induce a foreign adversary to back down in an international dispute. Crucially, the autocrat must weigh the benefits of forcing an enemy to back down against the potential cost of anti-foreign protests turning anti-regime. Ultimately, our model emphasizes that while nationalist protests can serve the autocrat as an audience cost, they are a blunt and soft tool that is only used under some circumstances.

Through this model, we contribute to our understanding of how authoritarian regimes can generate audience costs. Our explanation is most similar to work by Weiss (2013, 2014) in that we focus on nationalist protests. We build on Weiss and the broader literature on audience costs by linking it to the literature on censorship in political economy and build a microfoundation for the formation of audience costs which has received relatively little attention in the international relations literature (Slanchev 2006 is an exception).

Additionally, these findings extend our existing knowledge of media bias and censorship behavior in autocratic regimes. Prior explanations for increased media freedom have included increasing bureaucratic

accountability (Egorov et al. 2009) and a form of coup-proofing and protection from other elites (Boleslavsky et al. 2019; Hollyer et al. 2019). We argue that autocrats may reduce bias to generate nationalist protests and introduce risk to regime survival in order to send a credible signal of resolve during an international dispute.

We believe future research can go in several directions. First, more work, both theoretical and empirical, can be done on the relationship between authoritarian regimes, media policy, and audience costs. The model could be extended to include an explicit model for the domestic media market. Increasing media competition between both private and public media outlets may have ambiguous effects on the equilibrium level of bias (Guo 2020). Furthermore, the model could be extended to incorporate imperfect commitment to a media policy.

We also believe future research could consider a model where the foreign adversary can learn about the autocrat's and his citizens' beliefs via the choice of media policy. For example, the autocrat may have incentives to overly bias the media to signal hawkishness to the foreign enemy. Or she may want the enemy to receive a more dovish signal to de-escalate tensions at the risk of generating larger nationalist protest.

Additionally, this model informs future empirical analysis of censorship during interstate crisis. Some studies have just begun to accumulate descriptive statistics on this topic (e.g. Cairns and Carlson 2016; King et al. 2013). Our model can begin to make sense of and extend these types of studies and make predictions about how censorship policy might vary across crises.

A Proofs of Results in the Main Body

A.1 Proof of Proposition 1

Proof. The autocrat faces the following optimization problem

$$\max_{\beta(0) \in [0,1]} (\phi + (1 - \phi)\beta(0)) \left[\psi(z)(\phi\mu_H + (1 - \phi)\mu_L) + (1 - \psi(z))\phi(p\mu_H - k) - \theta(z)\pi \right] + (1 - \phi)(1 - \beta(0)) \left[\phi(p\mu_H - k) \right].$$

This problem has the following FOC with respect to $\beta(0)$

$$(1 - \phi) \left[\psi(z)[(\phi\mu_H + (1 - \phi)\mu_L) - \phi(p\mu_H - k)] - \theta(z)\pi \right] + (\phi + (1 - \phi)\beta(0))z' \left[\psi'(z)[(\phi\mu_H + (1 - \phi)\mu_L) - \phi(p\mu_H - k)] - \theta'(z)\pi \right] \begin{cases} < 0 & \text{if } \beta(0) = 0 \\ = 0 & \text{if } \beta(0) \in (0,1) \\ > 0 & \text{if } \beta(0) = 1, \end{cases}$$

where $\theta'(z)$, $\psi'(z)$, and z' denote the partial derivatives of $\theta(z)$, $\psi(z)$, and z with respect to $\beta(0)$.

Consider first the case where $\beta(0)^* \in (0,1)$, which produces expected protests of $z^* \in \left(0, \frac{\phi - \lambda}{1 - \bar{q}}\right)$.

Substituting the derivative of z , $-\frac{(\phi - \lambda)(1 - \phi)}{(1 - \bar{q})(\phi + (1 - \phi)\beta(0))^2}$, we rearrange to obtain

$$\beta(0)^* = \frac{(\phi - \lambda) \left[\psi'(z^*)\Sigma - \theta'(z^*)\pi \right] - \phi(1 - \bar{q}) \left[\psi(z^*)\Sigma - \theta(z^*)\pi \right]}{(1 - \phi)(1 - \bar{q}) \left[\psi(z^*)\Sigma - \theta(z^*)\pi \right]}.$$

Rearranging the expression for $\beta(0)^*$, we see that it is only in the interval $(0,1)$ when

$$\frac{\phi(1 - \bar{q})}{\phi - \lambda} < \frac{\psi'(z^*)\Sigma - \theta'(z^*)\pi}{\psi(z^*)\Sigma - \theta(z^*)\pi} < \frac{1 - \bar{q}}{\phi - \lambda}.$$

Taking the second derivative of the objective function yields

$$2(1-\phi)z' \left[\psi'(z)\Sigma - \theta'(z)\pi \right] + (\phi + (1-\phi)\beta(0))z'' \left[\psi'(z)\Sigma - \theta'(z)\pi \right] \\ + (\phi + (1-\phi)\beta(0))z'z'' \left[\psi''(z)\Sigma - \theta''(z)\pi \right],$$

where $\theta''(z)$, $\psi''(z)$, and z'' are the second partial derivatives of $\theta(z)$, $\psi(z)$ and z . And where $z'' = -\frac{2(\phi-1)(\lambda-\phi)}{(\bar{q}-1)((\phi-1)\beta(0)-\phi)^3}$, which is negative for all $\beta(0) \geq 0$. Recall that $\psi(z)$ is concave and $\theta(z)$ is convex, which implies that $\psi''(z)\Sigma - \theta''(z)\pi < 0$ for all z . Additionally, for $\beta(0) > 0$, $\psi'(z^*)\Sigma - \theta'(z^*)\pi$ must be positive. This means that the second derivative of the objective function is negative for all $\beta(0) > 0$. This implies the objective function is concave on the domain $(0,1)$. Therefore, when $\beta(0)^* \in (0,1)$, a global optimum is obtained and the autocrat's optimal strategy is to choose a level of bias equal to $\beta(0)^*$.

Now consider the case where $\beta(0) = 1$. Then no citizens watch the news because

$$\bar{q} > \phi > \phi - \lambda = \frac{\phi - \lambda}{\phi + \beta(0)(1 - \phi)},$$

which means no citizens protest. Then the expected utility of the autocrat is

$$\phi(p\mu_H - k).$$

Additionally, consider the case where $\beta(0) = 0$ and the autocrat generates expected protests of size $\frac{\phi-\lambda-\bar{q}}{1-\bar{q}}$, which is the maximum size of protest the autocrat can generate. Then the autocrat's expected utility is

$$\phi \left[\psi \left(\frac{\frac{\phi-\lambda}{\phi} - \bar{q}}{1-\bar{q}} \right) (\phi\mu_H + (1-\phi)\mu_L) \right. \\ \left. + (1 - \psi \left(\frac{\frac{\phi-\lambda}{\phi} - \bar{q}}{1-\bar{q}} \right)) \phi(p\mu_H - k) - \theta \left(\frac{\frac{\phi-\lambda}{\phi} - \bar{q}}{1-\bar{q}} \right) \pi \right] \\ + (1-\phi) \left[\phi(p\mu_H - k) \right].$$

Then, when an interior solution does not exist, the autocrat will receive a higher expected utility from setting $\beta(0) = 0$ relative to setting $\beta(0) = 1$ when

$$\psi\left(\frac{\frac{\phi-\lambda}{\phi} - \bar{q}}{1 - \bar{q}}\right)\Sigma > \theta\left(\frac{\frac{\phi-\lambda}{\phi} - \bar{q}}{1 - \bar{q}}\right)\pi,$$

will receive a higher expected utility from setting $\beta(0) = 1$ relative to $\beta(0) = 0$ when

$$\psi\left(\frac{\frac{\phi-\lambda}{\phi} - \bar{q}}{1 - \bar{q}}\right)\Sigma < \theta\left(\frac{\frac{\phi-\lambda}{\phi} - \bar{q}}{1 - \bar{q}}\right)\pi,$$

and will be indifferent between the two when

$$\psi\left(\frac{\frac{\phi-\lambda}{\phi} - \bar{q}}{1 - \bar{q}}\right)\Sigma = \theta\left(\frac{\frac{\phi-\lambda}{\phi} - \bar{q}}{1 - \bar{q}}\right)\pi.$$

■

A.2 Proof of Corollary 1

Proof. This follows from rearranging the partial derivative of the objective function with respect to $\beta(0)$. ■

A.3 Proof of Proposition 2

Proof. The four points in the proposition follow directly from the Implicit Function Theorem.

i. Note that the implicit function theorem implies that:

$$\frac{\partial\beta(0)}{\partial\lambda} = -\frac{F_\lambda}{F_{\beta(0)}}$$

Where $F = U'_A$ and F_λ and $F_{\beta(0)}$ are the partial derivatives of U'_A with respect to λ and $\beta(0)$, respectively. Then $F_{\beta(0)} = U''_A$ which was shown to be negative in the proof of proposition 1. Therefore, it is sufficient to show that $F_\lambda < 0$. This is true when

$$(\psi'(z)\Sigma - \theta'(z)\pi)\left(\frac{\partial(P(\hat{s}=1)z')}{\partial\lambda} + (1-\phi)\frac{\partial z}{\partial\lambda}\right) + P(\hat{s}=1)z'\frac{\partial z}{\partial\lambda}(\psi''(z)\Sigma - \theta''(z)\pi) < 0$$

This is always true because $\frac{\partial(P(\hat{s}=1)z')}{\partial\lambda} + (1-\phi)\frac{\partial z}{\partial\lambda} = 0$ and $P(\hat{s}=1)z'\frac{\partial z}{\partial\lambda}(\psi''(z)\Sigma - \theta''(z)\pi) < 0$.

ii. It is sufficient to show that $F_{\bar{q}} < 0$, which is true when

$$(\psi'(z)\Sigma - \theta'(z)\pi)\left(\frac{\partial(P(\hat{s}=1)z')}{\partial\bar{q}} + (1-\phi)\frac{\partial z}{\partial\bar{q}}\right) + P(\hat{s}=1)z'\frac{\partial z}{\partial\bar{q}}(\psi''(z)\Sigma - \theta''(z)\pi) < 0$$

Which is true for all parameters.

iii. It is sufficient to show $F_{\Sigma} < 0$. And $F_{\Sigma} = P(\hat{s}=1)z'\psi'(z) + (1-\phi)\psi(z)$ which is negative when $\frac{(1-\bar{q})P(\hat{s}=1)}{\phi-\lambda} < \frac{\psi'(z)}{\psi(z)}$. This is always true because

$$\frac{(1-\bar{q})P(\hat{s}=1)}{\phi-\lambda} < \frac{(1-\bar{q})}{\phi-\lambda} < \frac{\psi'(z)\Sigma - \theta'(z)\pi}{\psi(z)\Sigma - \theta(z)\pi} < \frac{\psi'(z)}{\psi(z)}$$

Therefore, we get $\frac{\partial\beta(0)}{\partial\Sigma} < 0$ for all $\beta(0)$.

iv. Similarly, we can use the implicit function theorem to find $\frac{\partial\beta(0)}{\partial\pi}$. It is sufficient to show that $F_{\pi} = -P(\hat{s}=1)z'\theta'(z) - (1-\phi)\theta(z) > 0$ which occurs when $\frac{\theta'(z)}{\theta(z)} > \frac{(1-\bar{q})P(\hat{s}=1)}{\phi-\lambda}$, which is always true because $\frac{\theta'(z)}{\theta(z)} > \frac{\psi'(z)}{\psi(z)}$ for all z .

■

A.4 Proof of Corollary 2

Proof. This follows from Proposition 2 and the partial derivatives of Σ .

- i. $\frac{\partial\Sigma}{\partial\mu_H} = \phi - \phi p$, which is positive for all parameters.
- ii. $\frac{\partial\Sigma}{\partial\mu_L} = 1 - \phi$, which is positive for all parameters.
- iii. $\frac{\partial\Sigma}{\partial p} = -\phi\mu_H$, which is negative for all parameters.
- iv. $\frac{\partial\Sigma}{\partial k} = -\phi$, which is positive for all parameters.

■

A.5 Proof of Lemma 1

Proof. We know that $\psi(z)$, $\theta(z)$, and $P(\hat{s} = 1)z' = -\frac{(\phi-\lambda)(1-\phi)}{(1-q)P(\hat{s}=1)}$ are all decreasing in $\beta(0)$. And because $\psi(z)$ is increasing and concave in z , we know that $\psi'(z)$ is decreasing for $\beta(0) \in [0,1]$, while $\psi(z)$ is increasing, resulting in the expected marginal benefit increasing over $\beta(0) \in [0,1]$. Likewise, $\theta(z)$ is increasing and convex in z , therefore $\theta'(z)$ is increasing for $\beta(0) \in [0,1]$, resulting in the expected marginal cost decreasing over $\beta(0) \in [0,1]$. ■

A.6 Proof of Proposition 3

Proof.

- i. Increasing the marginal probability that the enemy backs down, $P(\hat{s} = 1)z'\psi'(z) + (1 - \phi)\psi(z)$, for $\beta(0) \in (0,1)$ will shift the function in the right hand side of Corollary 1 downwards for all values of $\beta(0)$ resulting in an intersection with $\frac{\Sigma}{\pi}$ that is lower, resulting in a lower $\beta(0)^*$.
- ii. Increasing the marginal probability that protests turn anti-regime, $P(\hat{s} = 1)z'\theta'(z) + (1 - \phi)\theta(z)$, for $\beta(0) \in (0,1)$ will shift the function in the right hand side of Corollary 1 upwards for all values of $\beta(0)$ resulting in an intersection with $\frac{\Sigma}{\pi}$ that is higher, resulting in a higher $\beta(0)^*$.

■

Notes

1. For examples of work that argue autocrats struggle to generate audience costs see Fearon (1994), Stasavage (2007), Schultz and Weingast (2003).
2. See Gentzkow and Shapiro (2006) and Kamenica and Gentzkow (2011) for a discussion of Bayesian persuasion.
3. Throughout the paper, we consider the terms “nationalist” and “anti-foreign” protests to be interchangeable.
4. Specifically, Weiss (2013) notes four instances where nationalist protests were instrumental in leader removal in authoritarian regimes: Hungary 1956; Iran 1979; Madagascar 1972; Azerbaijan 1992. More recently, nationalist protests in Iran turned into anti-regime protests after the regime admitted to mistakenly shooting down a plane departing for Ukraine in January 2020; a nationwide crackdown soon followed.
5. Additional experimental evidence suggests that for the Chinese government, backing down or issuing empty threats can have negative effects on regime approval. See Li and Chen (2020), Weiss and Dafoe (2019), and Quek and Johnston (2018).
6. See Clines (1992)
7. See New York Times (1992)
8. See DellaVigna and Gentzkow (2010), Prat and Stromberg (2013), and Stromberg (2015) for reviews of the literature on mass media and its effect on politics.
9. For more discussion of Bayesian persuasion more broadly, see Kamenica and Gentzkow (2011) and Kamenica (2019).
10. This follows from the assumption that while the autocrat has control of the media, they still must delegate responsibility for reporting the news to reporters and editors who ultimately decide what makes it on the news. See Gehlbach and Sonin (2014) and Duggan and Martinelli (2011) for a similar approach
11. For simplicity, we often refer to the autocrat as having a “high” or “low” type.
12. The assumption that the autocrat does not know their own valuation with certainty at the start of the game reflects the fact that when the game starts the autocrat does not know all of the facts regarding the status of the conflict. Therefore, she does not know with certainty how much she values the outcome until time passes and she learns more.
13. See Gentzkow, Shapiro, and Sinkinson (2011) and Chiang and Knight (2011).
14. We assume that $\psi(x)$ is concave to capture the intuition that there are diminishing returns to the protests as they increase in size. After some point, the addition of the marginal protester does little to convey more with respect to the autocrat’s resolve.
15. We assume that the probability that anti-foreign protests become anti-regime protests is convex to reflect the discussion in Section 2.2. As protests increase in size they may pass a tipping point after which they rapidly become more costly because they quickly balloon in size or lead to elite schisms.
16. In line with Judd (1985), we assume that the law of large numbers hold and that all paths are measurable with respect to the uniform distribution.

17. The simplifying assumption can also be found in the literature, such as Guo (2020).

References

- Barber IV, B., & Miller, C. (2019). Propaganda and combat motivation: Radio broadcasts and German soldiers' performance in World War II. *World Politics*, 71(3), 457–502.
- Baum, M. A., & Potter, P. B. (2008). The relationships between mass media, public opinion, and foreign policy: Toward a theoretical synthesis. *Annual Review of Political Science*, 11, 39–65.
- Baum, M. A., & Potter, P. B. (2015). *War and democratic constraint: How the public influences foreign policy*. Princeton University Press.
- Boleslavsky, R., Shadmehr, M., & Sonin, K. (2021). Media freedom in the shadow of a coup. *Journal of the European Economic Association*, 19(3), 1782–1815.
- Cairns, C., & Carlson, A. (2016). Real-world islands in a social media sea: Nationalism and censorship on weibo during the 2012 Diaoyu/Senkaku crisis. *The China Quarterly*, 225, 23–49.
- Cairns, C., & Plantan, E. (2016). Signaling through the haze: Strategic (non)-censorship of online discussion of air pollution in China. *Working Paper*.
- Chiang, C.-F., & Knight, B. (2011). Media bias and influence: Evidence from newspaper endorsements. *The Review of economic studies*, 78(3), 795–820.
- Clines, F. X. (1992). Angry azerbaijanis impel chief to quit. *The New York Times*, 3–3. <https://www.nytimes.com/1992/03/07/world/angry-azerbaijanis-impel-chief-to-quit.html?searchResultPosition=4>
- De Mesquita, B. B., Smith, A., Siverson, R. M., & Morrow, J. D. (2003). *The logic of political survival*. MIT Press.
- DellaVigna, S., Enikolopov, R., Mironova, V., Petrova, M., & Zhuravskaya, E. (2014). Cross-border media and nationalism: Evidence from Serbian radio in Croatia. *American Economic Journal: Applied Economics*, 6(3), 103–32.
- DellaVigna, S., & Gentzkow, M. (2010). Persuasion: Empirical evidence. *Annual Review of Economics*, 2(1), 643–669.
- Ding, W. (2018). Propaganda, patriotic protest, and diplomatic persuasion. *Working Paper*.
- Duggan, J., & Martinelli, C. (2011). A spatial theory of media slant and voter choice. *The Review of Economic Studies*, 78(2), 640–666.
- Egorov, G., Guriev, S., & Sonin, K. (2009). Why resource-poor dictators allow freer media: A theory and evidence from panel data. *American Political Science Review*, 103(4), 645–668.
- Enikolopov, R., Petrova, M., & Zhuravskaya, E. (2011). Media and political persuasion: Evidence from Russia. *American Economic Review*, 101(7), 3253–85.

- Eyerman, J., & Hart Jr, R. A. (1996). An empirical test of the audience cost proposition: Democracy speaks louder than words. *Journal of Conflict Resolution*, *40*(4), 597–616.
- Fearon, J. D. (1994). Domestic political audiences and the escalation of international disputes. *American Political Science Review*, *88*(3), 577–592.
- Fearon, J. D. (1997). Signaling foreign policy interests: Tying hands versus sinking costs. *Journal of Conflict Resolution*, *41*(1), 68–90.
- Feddersen, T., & Pesendorfer, W. (1998). Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting. *American Political Science Review*, 23–35.
- Gehlbach, S., & Sonin, K. (2014). Government control of the media. *Journal of Public Economics*, *118*, 163–171.
- Gelpi, C. F., & Griesdorf, M. (2001). Winners or losers? Democracies in international crisis, 1918–94. *American Political Science Review*, *95*(3), 633–647.
- Gentzkow, M., & Shapiro, J. M. (2006). Media bias and reputation. *Journal of Political Economy*, *114*(2), 280–316.
- Gentzkow, M., Shapiro, J. M., & Sinkinson, M. (2011). The effect of newspaper entry and exit on electoral politics. *American Economic Review*, *101*(7), 2980–3018.
- Guo, X. (2020). Media censorship backfire. *Working Paper*.
- Hollyer, J. R., Rosendorff, B. P., & Vreeland, J. R. (2019). Why do autocrats disclose? Economic transparency and inter-elite politics in the shadow of mass unrest. *Journal of Conflict Resolution*, *63*(6), 1488–1516.
- Ipek, P. (2009). Azerbaijan’s foreign policy and challenges for energy security. *The Middle East Journal*, *63*(2), 227–239.
- Judd, K. (1985). The law of large numbers with a continuum of IID random variables. *Journal of Economic Theory*, *35*(1), 19–25.
- Kamenica, E. (2019). Bayesian persuasion and information design. *Annual Review of Economics*, *11*, 249–272.
- Kamenica, E., & Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, *101*(6), 2590–2615.
- King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *American Political Science Review*, *107*(2), 326–343.
- Kuran, T. (1991). Now out of never: The element of surprise in the East European revolution of 1989. *World Politics*, *44*(1), 7–48.
- Li, X., & Chen, D. (2020). Public opinion, international reputation, and audience costs in an authoritarian regime. *Forthcoming, Conflict Management and Peace Science*.

- Little, A. T. (2016). Communication technology and protest. *The Journal of Politics*, 78(1), 152–166.
- Lorentzen, P. (2014). China’s strategic censorship. *American Journal of Political Science*, 58(2), 402–414.
- Luo, Z., & Rozenas, A. (2018). Strategies of election rigging: Trade-offs, determinants, and consequences. *Quarterly Journal of Political Science*, 13(1), 1–28.
- Ma, X. (2020). Playing with firecrackers: When and why nationalist protests at home affect crisis bargaining abroad. *Working Paper*.
- Manacorda, M., & Tesei, A. (2020). Liberation technology: Mobile phones and political mobilization in Africa. *Econometrica*, 88(2), 533–567.
- Mattingly, D., & Yao, E. (2020). How propaganda manipulates emotion to fuel nationalism: Experimental evidence from China. *Working Paper*.
- O’Donnell, G., & Schmitter, P. C. (1986). *Transitions from authoritarian rule: Tentative conclusions about uncertain democracies*. Johns Hopkins University Press.
- Pan, Z., & Kosicki, G. M. (1994). Voters’ reasoning processes and media influences during the persian gulf war. *Political Behavior*, 16(1), 117–156.
- Potter, P. B., & Baum, M. A. (2010). Democratic peace, domestic audience costs, and political communication. *Political Communication*, 27(4), 453–470.
- Potter, P. B., & Baum, M. A. (2014). Looking for audience costs in all the wrong places: Electoral institutions, media access, and democratic constraint. *The Journal of Politics*, 76(1), 167–181.
- Prat, A., & Strömberg, D. (2013). The political economy of mass media. *Advances in Economics and Econometrics*, 2, 135.
- Przeworski, A. (1991). *Democracy and the market: Political and economic reforms in Eastern Europe and Latin America*. Cambridge University Press.
- Qin, B., Strömberg, D., & Wu, Y. (2019). Social media, information networks, and protests in China. *Working Paper*.
- Qin, B., Strömberg, D., & Wu, Y. (2017). Why does China allow freer social media? Protests versus surveillance and propaganda. *Journal of Economic Perspectives*, 31(1), 117–40.
- Quek, K., & Johnston, A. I. (2018). Can China back down? Crisis de-escalation in the shadow of popular opposition. *International Security*, 42(3), 7–36.
- Schultz, K. A. (2001). *Democracy and coercive diplomacy* (Vol. 76). Cambridge University Press.
- Schultz, K. A., & Weingast, B. R. (2003). The democratic advantage: Institutional foundations of financial power in international competition. *International Organization*, 57(1), 3–42.
- Shadmehr, M., & Bernhardt, D. (2015). State censorship. *American Economic Journal: Microeconomics*, 7(2), 280–307.

- Slantchev, B. L. (2006). Politicians, the media, and domestic audience costs. *International Studies Quarterly*, 50(2), 445–477.
- Soroka, S. N. (2003). Media, public opinion, and foreign policy. *Harvard International Journal of Press/Politics*, 8(1), 27–48.
- Stasavage, D. (2007). Cities, constitutions, and sovereign borrowing in Europe, 1274-1785. *International Organization*, 489–525.
- Strömberg, D. (2015). Media and politics. *Annual Review of Economics*, 7, 173–205.
- Times, T. N. Y. (1992). Revolt in capital of azerbaijan. *The New York Times*, 4–4. <https://www.nytimes.com/1992/05/16/world/revolt-in-capital-of-azerbaijan.html?searchResultPosition=1>
- Wallace, J. (2013). Cities, redistribution, and authoritarian regime survival. *The Journal of Politics*, 75(3), 632–645.
- Wallace, J. (2014). *Cities and stability: Urbanization, redistribution, and regime survival in China*. Oxford University Press.
- Weeks, J. L. (2008). Autocratic audience costs: Regime type and signaling resolve. *International Organization*, 62(1), 35–64.
- Weeks, J. L. (2012). Strongmen and straw men: Authoritarian regimes and the initiation of international conflict. *American Political Science Review*, 326–347.
- Weiss, J. C. (2013). Authoritarian signaling, mass audiences, and nationalist protest in China. *International Organization*, 67(1), 1–35.
- Weiss, J. C. (2014). *Powerful patriots: Nationalist protest in China's foreign relations*. Oxford University Press.
- Weiss, J. C., & Dafoe, A. (2019). Authoritarian audiences, rhetoric, and propaganda in international crises: Evidence from China. *International Studies Quarterly*, 63(4), 963–973.