British Journal of Political Science (2024), **54**, 313–338
doi:10.1017/S0007123423000182

British Journal of Political Science

ARTICLE

Identity Propaganda

Carlo M. Horz

Department of Political Science, Texas A&M University, College Station, TX 77845-4348, USA
Email: carlo.horz@tamu.edu

(Received 7 April 2022; revised 21 December 2022; accepted 4 April 2023; first published online 25 May 2023)

Abstract

Political elites often employ propaganda to affect the behavior of a particular social group by altering its members' social identities. The empirical literature has demonstrated that this kind of 'identity propaganda' is generally effective at mobilizing citizens. However, while the consequences of being exposed to propaganda depend on its content, we know little about which factors shape propaganda content. To gain insight into the determinants of propaganda content, I analyze a game-theoretic model where a political elite proposes a new identity norm, and citizens affirm or reject it. I demonstrate that, in equilibrium, the propagandist exploits his agenda-setting power to design effective identity norms. I also show that more demanding identity norms can emerge when citizens' mobilization costs are higher, or the propagandist can cheaply allocate material incentives. By contrast, the nature of strategic interaction among citizens has an ambiguous effect on identity norms.

Keywords: identity; propaganda; game theory

Political elites frequently make propagandistic statements about the content of citizens' social identities; for example, that a given social identity implies particular loyalties to the leader or special enmity against an out-group (Brass 1997; Eifert, Miguel, and Posner 2010; Horowitz 1985). Empirical studies on several countries and political contexts have shown that such identity propaganda can powerfully affect citizens' attitudes and behaviors. For example, exposure to ISIS social media propaganda broadly increases support for the group, but violent messages are less persuasive (Mitts, Phillips, and Walter 2022).[1] Similarly, exposure to Nazi radio broadcasts increased membership in the Nazi party, anti-Semitic acts, and denunciations shortly after the Nazi dictatorship began (Adena et al. 2015), as well as fighting efforts by German soldiers later on (Barber and Miller 2019). Finally, exposure to radio stations in Rwanda seems to have aided in the commission of genocide (Yanagizawa-Drott 2014) and the improvements in ethnic relations after that (Blouin and Mukand 2019).

From these examples, it is clear that identity propaganda's extremeness – in terms of content or practical exhortations – can vary widely across places and times, from calls for solidarity with the leader or the in-group to the encouragement of discrimination to the implementation of genocide.[2] The apparent effectiveness of extreme propaganda implies that it is important to understand the conditions under which leaders choose to air it. Unfortunately, our knowledge of the determinants of propaganda's content is limited.

[1]In general, ISIS and other Islamic terrorist groups argue that fighting against Western powers is every Muslim's duty, while more moderate groups define Jihad without invoking fighting at all (Askew and Helbardt 2012), which underscores the importance of content.

[2]For a case study on these themes, see Kershaw (2001) and Welch (2014).

Empirically, studying the determinants of propaganda content requires measuring it comparably across cases. Most empirical studies of identity propaganda focus on a specific source of propaganda in a particular country and event. A few scholars have attempted to compare propaganda aspects across countries and times, but they are not directly concerned with identity propaganda's content. Carter and Carter (2021) introduce a cross-national measure of propaganda based on newspaper data, focusing on leader valence, not citizens' identities. DellaVigna and Gentzkow (2010) compute the 'persuasion rate' for a number of studies, but this is a measure of effectiveness, not content.[3]

Theoretically, formal theories of propaganda and social identities usually focus on other quantities of interest. First, current formal-theoretic models of propaganda almost exclusively focus on propaganda concerning facts, such as a regime's economic performance, popularity, or military capacity (see, for example, Callander and Wilkie 2007; Chen and Xu 2017; Edmond 2013; Gehlbach and Sonin 2014; Huang 2015; Little 2017). These contributions (by design) cannot address propaganda about social identities – parameters for which no exogenous truth exists – such as what it means to be a social group member in a particular situation.[4] Second, the bulk of the theoretical work on endogenous social identities focuses either on the incentives of citizens to choose particular identities (Bénabou and Tirole 2011; Penn 2008; Sambanis and Shayo 2013; Schnakenberg 2014; Shayo 2009) or they assume propaganda is effective to study how propaganda affects leaders' other choices (for example, Dickson and Scheve 2006; Lehmann and Tyson 2022; Mukand and Rodrik 2018).

In this paper, I construct a behavioral game-theoretic model of identity propaganda to understand how the extremeness of identity propaganda might vary, taking into account both the incentives of elites to design effective propaganda and the incentives of citizens to affirm particular identities.[5] I focus on identity norms, which I define as behavior commonly understood to be prescribed by social identity; that is, it is an 'injunctive norm' (Reno, Cialdini, and Kallgren 1993).

I begin by analyzing a baseline model that provides a parsimonious account of endogenous identity propaganda content. The players of the game are a propagandist ('she') and a citizen ('he'). They belong to the same social group, and I assume the propagandist has the authority to propose a new identity norm. In the model, the propagandist's message is equal to the following exhortation: as a member of this social group, it is your duty to behave in this way; that is, to exert a particular amount of effort. Following the propagandist's exhortation, the citizen decides whether to affirm or reject this new norm. If he affirms it, the propagandist's proposed norm becomes the citizen's active norm; if he rejects it, the citizen's active norm remains the existing identity norm, which is an exogenous parameter. Next, the citizen chooses how much effort to exert to increase the probability of some (binary) political outcome; for example, a regime change, winning an election, defeating another group in a conflict, or providing a community good. I assume that the citizen derives some material benefit from bringing about an outcome, which incentivizes him to invest in effort, but effort is costly. I refer to the parameter indexing the costs of effort as mobilization costs. In addition, the citizen also cares about behaving in accordance with his active social identity; in particular, he faces an additional loss function around the level of effort prescribed by the identity norm.[6] The propagandist, by contrast, cares only about maximizing the probability of a certain political outcome.

[3]The persuasion rate estimates 'the percentage of receivers that changed their behavior among those that receive a message and are not already persuaded' (DellaVigna and Gentzkow 2010, 645). For a critique of this measure, see Jun and Lee (2018).

[4]Some formal-theoretic work incorporates identity propaganda by treating it as fact-based (Baliga and Sjöström 2012; Glaeser 2005).

[5]Following existing work (for example, Jowett and O'Donnell 2015), I define identity propaganda as communication about social identity that aims to change the audience's perceptions and improve the sender's welfare.

[6]As I explain in detail below, this specification is consistent with Akerlof and Kranton's (2000) conceptualization of identity. In addition, Akerlof and Kranton (2005) consider a linear loss function in their treatment of the effects of identity in principal-agent contracting environments. Below, I provide a more detailed comparison with Akerlof and Kranton (2005).

In equilibrium, the citizen's level of effort depends on the material stakes (the difference in utility between the preferred and the undesired outcome), his mobilization costs, and the active identity norm. I demonstrate that the more demanding the citizen's identity norm, the higher the effort exerted. Thus, the propagandist wishes to design the most demanding norm possible but has to consider the citizen's incentives to affirm a particular proposed identity norm. Specifically, for some proposed identity norms, there is a tension between the citizen's material and identity concerns – satisfying material concerns is not the same as satisfying identity concerns. I show that citizens are willing to accept new identity content if the tension between these concerns is reduced. Formally, I demonstrate that there is a unique identity norm that eliminates the tension between material and identity concerns, and identity norms are accepted if they are close to this 'optimal' norm. The propagandist anticipates these considerations by the citizen and chooses the most demanding identity norm that the citizen still accepts.

While deliberately sparse, the baseline model delivers several important insights. First, in equilibrium, using identity propaganda is an effective strategy; relative to a world without leader communication, the citizen's level of effort is higher. This is consistent with the recent quantitative literature on propaganda (discussed above) that documents positive causal effects on citizens' choices (Adena et al. 2015; Barber and Miller 2019; Bleck and Michelitch 2017; Blouin and Mukand 2019; DellaVigna et al. 2014; Enikolopov, Petrova, and Zhuravskaya 2011; Mitts, Phillips, and Walter 2022; Peisakhin and Rozenas 2018; Yanagizawa-Drott 2014).[7]

Second, propaganda's content is more demanding or extreme when the citizen's stakes are higher, when mobilization costs are lower, or when existing norms are relatively less demanding. These predictions support existing studies that, while not directly concerned with propaganda content, feature findings consistent with the model. In particular, I discuss the baseline model's implications concerning communication and behavior during the COVID-19 pandemic, inter-group conflicts, and national crises. My model points out that norms were inconsistent with many new public health guidelines at the beginning of the pandemic, which may have enabled effective, norm-based leader communication. My model also provides insights into the dynamics of inter-group conflicts, explaining why citizens' material incentives (that is, their stakes) play a crucial role in explaining effective identity propaganda in ethnic conflicts. Finally, I explain why threat perception can increase leader support in 'rally-'round-the-flag' studies (Kobayashi and Katagiri 2018).

I then enrich the baseline model to scrutinize the effect of more realistic features of the political environment on identity propaganda's content. I focus on three factors that previous work has shown to be relevant for leader communication and citizen behavior: the extent to which there is preference heterogeneity among citizens, the nature of strategic interaction when mobilizing to achieve a political objective, and the presence of material incentives (Alonso and Câmara 2016; Dickson 2010; Landa and Tyson 2017). These features are difficult to measure; hence, conducting an empirical test that directly measures their effect on propaganda content is challenging. However, these important features vary across study contexts. By analyzing their effect on equilibrium choices, I help explain existing empirical results and point to sources of variation in the design of propaganda. I also discuss empirical findings that resonate with the theoretical predictions.

I first consider the effect of preference heterogeneity across citizens, analyzing a situation where there are two (groups of) citizens – one of which cares more about the political outcome. I show that the citizen with more at stake in the outcome is more ready to affirm new identity norms than the citizen with less at stake. As a consequence, the propagandist faces the following choices, (1) she can propose a more moderate identity norm that is accepted by both group

[7]The analysis is also consistent with the evidence presented in the literature on injunctive norms (for example, Hallsworth et al. 2017; Reno, Cialdini, and Kallgren 1993) and the experimental literature on social identities (for example, Chen and Li 2009; Landa and Duell 2015).

members ('universal norm') or (2) she can propose a more demanding identity norm that is only accepted by one member ('divisive norm'). I find that proposing a more demanding identity norm is more attractive the larger the size of the faction that has more at stake in the outcome, the greater the mobilization costs, and the more demanding the pre-existing identity norm. These results have two important implications. First, while the size of the faction that has more at stake has a positive effect on the emergence of divisive norms, depending on parameter values, choosing a divisive norm can be optimal even if most members are moderates. Second, an increase in mobilization costs (or the existing identity norm) has a non-monotone (hence: ambiguous) effect on the emergence of demanding identity norms. On the one hand, an increase in either makes citizens less likely to affirm demanding identity norms. On the other hand, an increase in either incentivizes the propagandist to choose the divisive, partially affirmed identity norm. Focusing on the latter effect, existing scholarship finds some support for it; populist candidates' rhetoric seems more extreme when attempting to mobilize relatively marginalized voters (Anduiza, Guinjoan, and Rico 2019; Cheeseman and Larmer 2015; de la Torre 2017; McCoy and Somer 2019), often encouraging them to 'fight like hell' in the face of high obstacles (NPR 2021). In general, however, my analysis implies that future empirical scholarship scrutinizing the effect of mobilization costs on identity norms can find a positive or negative relationship, depending on the exact comparison.

I next analyze the effect of strategic interaction among citizens with identical preferences. In some situations, citizens' effort choices are strategic complements, i.e., citizens' incentives to choose high levels of effort are higher when other citizens also choose a high level of effort. In other situations, citizens' effort choices are strategic substitutes, i.e., citizens' incentives to choose high levels of effort are higher when other citizens choose a low level of effort. I adapt the baseline model to take such incentives into account, showing that for some proposed identity norms, strategic interaction induces multiple equilibria when the citizens choose to affirm newly proposed identity norms. Moreover, the nature of strategic interaction determines the nature of multiplicity. When effort levels are strategic complements, there are multiple symmetric equilibria, i.e., there is an equilibrium in which both citizens affirm but also an equilibrium in which neither citizen affirms. By contrast, when effort levels are strategic substitutes, there are multiple asymmetric equilibria, i.e., there are two equilibria in which exactly one citizen affirms while the other rejects the proposed identity norm. However, in both cases, citizens' expectations about identity change have a causal effect on identity norms. This finding speaks to two pieces of evidence. First, scholars have documented that seemingly similar organizations can have very different organizational norms (Carrillo and Gromb 1999; Gibbons 2010). Second, partly as a consequence, seemingly similar organizations can have very different success rates (Amenta et al. 2010). My finding explains why groups or organizations that initially look very similar in terms of material conditions can end up with very different identity norms – and, hence, long-run outcomes.

Finally, I study the effect of carrot-and-stick tactics (repression or cooptation). Suppose the propagandist can punish the citizen for bringing about a bad outcome or reward him for achieving a good outcome. I show that employing these material incentives increases citizens' proclivity to affirm more demanding identity norms. This, in turn, increases the propagandist's return from using such carrot-and-stick tactics; in equilibrium, this means that a higher level of such tactics is used than in a model without identity propaganda. Thus, there is a novel source of complementarity between material incentives and propaganda. This result provides one possible rationale for why regimes that invest heavily in propaganda are often the most repressive; for example, Nazi Germany and the Soviet Union (Gehlbach 2018). This model variation can also explain effective identity propaganda in cases where the citizen's material incentives seem completely opposed to the leader's; using material incentives can help align material interests, enabling effective identity propaganda.

Besides contributing to the theoretical and empirical literature on propaganda, my paper also adds to the literature on leadership (Ahlquist and Levi 2011) and authoritarian politics

(Gehlbach, Sonin, and Svolik 2016). In each literature, scholarship has shown that leaders employ communication to influence followers and that features such as the audience's heterogeneity (for example, Alonso and Câmara 2016), strategic interaction among citizens (for example, Dickson 2010; Torun and Myatt 2007), or the availability of material incentives (for example, Landa and Tyson 2017; Tyson and Smith 2018) are important factors for explaining variation in the efficacy and content of leader communication. My contribution is to extend existing analyses to a different kind of communication: identity propaganda.

The Baseline Model

Setup

The players of the game are a propagandist (an 'identity entrepreneur') denoted by P ('she') and a citizen ('he'). I assume the propagandist and the citizen belong to the same social group, for example, a nation, class, or religious group. Given the citizen's membership in the group, I sometimes refer to the citizen as a 'group member.' In the Appendix, I examine the cases in which several (identical) citizens form the audience for propaganda and in which the citizen is a member of several social groups.

The players' interests are built around a binary political outcome denoted by $y \in \{0, 1\}$. For example, $y = 1$ can be interpreted as winning a (civil or interstate) conflict, providing a good for the community, achieving regime change, or winning an election. For now, I assume the probability of outcome $y = 1$ is equal to the citizen's level of effort, denoted by $e \in [0, 1]$:

$$\Pr(y = 1|e) \equiv \gamma(e) = e. \tag{1}$$

For example, the citizen can supply a fighting effort to win a conflict, contribute by working to provide a good community, protest to topple a dictator, mobilize voters to win an election, or engage in social distancing to avoid a major outbreak of a disease.

The propagandist wishes to obtain the outcome $y = 1$ so her utility function is:

$$U_P = y.$$

The group member's utility function is more complex; I assume that citizens are concerned about material well-being while also having identity concerns. Thus, their utility function is given by:

$$U(e) = U^M(e) + U^I(e),$$

where U^M represents material concerns while U^I represents identity concerns. I assume that identity propaganda affects the shape of the identity function U^I. In principle, a propagandist could attempt to choose the entire function U^I. However, for tractability, in the main text, I focus on the following identity utility function:

$$U^I = -\frac{\alpha}{2}(e - \hat{e})^2. \tag{2}$$

where, α is a measure of salience, and the term $\hat{e}$ represents that identity norm, which I interpret as the level of effort prescribed by the citizen's identity. Given the political situation, this prescription states how the citizen should behave as a social group member. The term $\hat{e}$ can either be equal to $\hat{e}_E$ which is the existing identity norm (before the situation unfolds) or equal to $\hat{e}_P$ which is the identity norm championed by the propagandist. I describe the process by which the existing or the proposed norm is 'active' in the citizen's utility function below.

The functional form of U^I in expression 2 is convenient for several reasons: first, any behavior ($e \in [0, 1]$) can be advocated as being appropriate for a group member at a given time. Second, it

is also very tractable. Note, however, that the functional form implies that both 'too little' ($e < \hat{e}$) and 'too much' ($e > \hat{e}$) effort is costly.[8] I discuss this important assumption in more detail below.

Moreover, in principle, identity propaganda could affect both salience α and content $\hat{e}$. In the main text, I focus on endogenizing the identity norm $\hat{e}$. This makes sense when focusing on applications such as shaping citizens' behavior in a public health crisis or in inter-group conflicts when leaders directly communicate about appropriate behavior. However, in the Appendix, I consider variations in the identity portion of the citizen's utility function, the possibility of identity propaganda altering the citizen's social preferences, and the possibility that propaganda changes identity salience.[9] Finally, while it is important to emphasize that U^I is a pure loss function, this is inconsequential for the analysis, and adding a positive constant would yield identical results.[10]

Since the norm $\hat{e}$ prescribes a particular level of effort, I refer to norms that demand higher levels of effort as more demanding and norms that demand smaller levels of effort as less demanding. Given this definition of the identity norm, we can think of U^I as the psychological costs associated with deviating from the prescribed behavior $\hat{e}$ (Akerlof and Kranton 2000).

Second, there is a portion of the utility function based on material considerations:

$$U^M = yu(1) + (1-y)u(0) - \frac{c}{2}e^2,$$

where $u(y)$ is the utility of achieving outcome y. I assume the citizen is better off when the outcome is $y = 1$ relative to the situation in which the outcome is $y = 0$; that is, $u(1) > u(0)$. Furthermore, the costs of effort are quadratic, so the citizen faces the loss function $\frac{c}{2}e^2$. I refer to the parameter c as the citizen's mobilization costs. For example, these mobilization costs may be higher if the citizen attempts to overthrow an authoritarian regime and the regime is highly repressive. It may also represent opportunity costs, so in a turnout application, a higher level of c may refer to voters or activists who face relatively large opportunity costs and are, therefore, relatively 'marginalized' in the political system.

The following sequence of moves gives the process by which identity norm change may come about:

(1) The propagandist proposes $\hat{e}_P \in [0, 1]$.
(2) The group member chooses to affirm the norm $\hat{e}_P$, denoted by $t = 1$, or to keep his pre-existing identity norm $\hat{e}_E$, denoted by $t = 0$. Affirmation changes the citizen's identity payoff function to $U^I = -\frac{\alpha}{2}(e - \hat{e}_P)^2$ while not affirming it implies the existing loss function $U^I = -\frac{\alpha}{2}(e - \hat{e}_E)^2$.
(3) The group member chooses effort $e \in [0, 1]$, taking the identity norm as given.
(4) Nature chooses the outcome $y = 1$ or the outcome $y = 0$ according to expression 1.

It is important to emphasize that the model is not a cheap talk or a Bayesian Persuasion model. In contrast to those models, there is no uncertainty over a state of the world or a type here. Using the terminology introduced in DellaVigna and Gentzkow (2010), such models are belief-based, whereas my model is an instance of a preference-based persuasion model. Given this setup, I look for an equilibrium of the following form:

[8]The term $\hat{e}$ may be interpreted as the effort level corresponding to the group's prototype; hence, the norm 'may both encourage effort by some group members (those below the prototypical effort) and discourage effort among those above the norm' (Shayo 2020, 359).

[9]I consider both the case in which there is an increase in identity salience relative to the citizen's material interests and relative to the citizen's other social identities.

[10]As I explain below and in more detail in the Appendix, given the equilibrium concept employed here, what matters is the interaction of effort e and identity norm $\hat{e}$, not the 'scale' of the identity utility function U^I.

Definition 1. *An equilibrium is a tuple* $(e^*, t^*, \hat{e}_P^*)$ *such that:*

(1) *Optimal effort:* $e^*(\hat{e}) \in \arg\max U^M(e) + U^I(e, \hat{e})$.
(2) *Optimal affirmation:*
$t^*(e_P) = 1$ if $U^M(e^*(\hat{e}_P)) + U^I(e^*(\hat{e}_P), \hat{e}_P) \geq U^M(e^*(\hat{e}_E)) + U^I(e^*(\hat{e}_E), \hat{e}_E)$.
(3) *Optimal propaganda:* $\hat{e}_P^* \in \arg\max t^*(\hat{e}_P)y(e^*(\hat{e}_P)) + (1 - t^*(\hat{e}_P))y(e^*(\hat{e}_E))$.

This is simply a subgame perfect Nash equilibrium in which one choice affects one player's future payoff function. A similar definition (for simultaneous move games) is employed by Shayo (2009) and Sambanis and Shayo (2013). These assumptions imply that two conditions have to be satisfied for identity change. First, the propagandist must decide to propose an identity norm that differs from the existing identity norm; that is, $\hat{e}_P \neq \hat{e}_E$. Second, the citizen has to affirm this proposed identity norm $\hat{e}_P$ rather than sticking with the existing identity norm $\hat{e}_E$.

Discussion of Assumptions

Before proceeding to the analysis, I discuss the model's key assumptions. The existing literature on social identities has conceptualized identity concerns in various ways, ranging from social preferences such as spite, altruism, or group status concerns (Chen and Li 2009) to perceived distance to group ideal types (Shayo 2009) and prescribed actions (Akerlof and Kranton 2000). My conceptualization of identity draws heavily on the Akerlof-Kranton framework of identity (Akerlof and Kranton 2000; Akerlof and Kranton 2005; see also Dickson and Scheve 2006 and Dickson and Scheve 2010), which treats identities as social categories that are associated with particular (exogenous) behavioral prescriptions, which are denoted here by the identity norm $\hat{e}$.

Following prescribed identity norms can require materially costly behavior, but not following them is also costly and is associated with negative emotions such as anxiety, dread, or depression.[11] In Akerlof and Kranton (2005), this is modeled as a linear loss function. However, I employ a quadratic loss function, which improves tractability because it is differentiable everywhere. Note that either specification implies that effort levels lower than the prescribed action ($e < \hat{e}$) and effort levels that are higher than the prescribed action ($e > \hat{e}$) are (equally) psychologically costly. This is a helpful simplification in the present context, but it may not always be realistic. For example, the propagandist may air a norm such as 'You must do at least $\hat{e}$, but doing more is fine.' I briefly consider such possibilities in the Appendix and provide an analysis that relies on different functional forms U^I. I discuss the results below in the conclusion section.

As mentioned, Akerlof and Kranton (2005) also feature an effort choice and a loss function around a prescribed identity action. However, their substantive focus, and hence their analysis, is very different. Akerlof and Kranton (2005) focus on a contracting environment in which a principal offers a contract (that is, a wage payment contingent on an observed outcome) to influence the agent's effort. They focus on the contracting literature and the relationship between the optimal contract and exogenous identity prescriptions. By contrast, my focus is on the literature on propaganda and on endogenizing the ideal level of effort prescribed by the agent's identity. Finally, they do not scrutinize the role of citizens' preference heterogeneity, strategic interaction among agents, or repression.

In general, unlike existing applications of the Akerlof and Kranton (2000) framework (for example, Dickson and Scheve 2006; Dickson and Scheve 2010), I endogenize the identity norm to which the citizen subscribes through a hierarchical process that can be interpreted as communication. On a formal level, it is essentially a complete information bargaining game in which the propagandist (the 'agenda setter') has all the bargaining power and the citizen's

[11] In a broader interpretation of this portion of the model, these psychological costs interact with material costs because of peer punishment and ostracism (see Kalin and Sambanis 2018, 242).

'outside option' is to stick to the existing identity norm (the citizen has 'veto rights'). This setup requires two important assumptions: first, I assume that the citizen can choose to change his identity norm if a different norm is offered; second, I assume that the citizen cannot propose a new norm to himself and then affirm it.

With regard to the first assumption, substantial research in the ethnic politics literature supports the claim that individuals can choose to alter the content of their social identities (for reviews of the evidence, see Akerlof and Kranton 2000; Shayo 2009). The model's insights are also robust, allowing the citizen to only alter his identity with a certain probability. In this case, all choices must be adjusted by the probability with which this occurs. Finally, to ensure that the model is consistent with existing behavioral game-theoretic work on endogenous identity formation, I assume that, while the citizen's psychological payoff is affected by his choice of a new social identity – he feels 'bad' if he does not live up to his identity prescriptions – his material payoff is not.

I make the second assumption for several related reasons. First, it is clear that in the real world, changing individual identity norms is a complex process in which many actors influence the outcome. The exact role each plays in determining the new norm is difficult to determine. While individuals have some degree of choice in their identity, other actors (members of the same group and members of other groups) and the extent to which immutable characteristics define identities also play a role (see Huddy 2001, 140–1, for a discussion). Second, I consider the simplest possible setup incorporating such an agreement – in the form of elite- and mass-influence – which the propagandist proposes and the citizen affirms or rejects. This conceptualization is also consistent with extensive empirical evidence on elites' agenda-setting power (see, for example, Fearon and Laitin 2000). Third, proposing new identity norms is also costly since a group's history, myths, and values have to be linked with a specific action in a convincing way. In the Appendix, I demonstrate that my analysis can be interpreted as applying to those citizens for whom proposing their identity norms is too costly.

Analysis

Equilibrium

Consistent with the equilibrium concept outlined above, I consider first the citizen's choice of effort, taking the active identity norm as given. Then, the group member solves the following optimization problem:

$$\max_{e\in[0,1]} \; e\Delta - \frac{c}{2}e^2 - \frac{\alpha}{2}(e-\hat{e})^2$$

where $\Delta \equiv u(1) - u(0)$ is defined as the citizen's stakes in the political situation. This yields the following optimal effort choice:

$$e^*(\hat{e}) = \begin{cases} \dfrac{\Delta + \alpha\hat{e}}{\alpha + c} & \text{if } \dfrac{\Delta + \alpha\hat{e}}{\alpha + c} < 1 \\ 1 & \text{otherwise.} \end{cases} \tag{3}$$

Intuitively, when the effort choice is interior, then the optimal effort choice is higher the larger the citizen's stakes in the situation and the lower the mobilization costs are. The effect of the salience term α on effort is ambiguous: it depends on the relative sizes of the current identity norm, the stakes, and the mobilization costs.[12] Most importantly, the optimal effort is higher the larger the citizen's identity norm because the citizen wishes to avoid the psychological loss associated

[12]Formally, $\frac{\partial e^*}{\partial \alpha} = \frac{c\hat{e}-\Delta}{(\alpha+c)^2}$ which can be positive or negative.

with deviating from an identity norm that demands higher effort. To keep matters simple, I assume throughout that effort is interior for all possible identity norms, which is implied by $c > \Delta$.

Now consider the citizen's decision to affirm or not affirm a new identity norm. To make this decision, he must determine the consequences of each action – and then evaluate which identity norms are 'acceptable.' Consider the equilibrium utility, or value, of an arbitrary identity norm $\hat{e}$:

$$V(\hat{e}) \equiv U(e^*(\hat{e}), \hat{e}) = e^*(\hat{e})\Delta - \frac{c}{2}e^*(\hat{e})^2 - \frac{\alpha}{2}(e^*(\hat{e}) - \hat{e})^2 + u(0)$$

I can now formally define what it means for an identity norm to be 'optimal:'

Definition 2. *The citizen's* optimal identity norm *is the identity norm that maximizes the equilibrium utility* $V(\hat{e})$*:*

$$\hat{e}_{\text{opt}} \in \underset{\hat{e}\in[0,1]}{\arg\max}\ V(\hat{e})$$

An important goal of the analysis is to contrast the citizen's optimal norm with the norm emerging in equilibrium. In order to make progress on characterizing the equilibrium identity norm, consider the change in the citizen's equilibrium utility as the identity norm $\hat{e}$ becomes more demanding:

$$\frac{\partial V}{\partial \hat{e}} = \frac{\partial e^*}{\partial \hat{e}} \underbrace{[\Delta - ce^*(\hat{e}) - \alpha(e^*(\hat{e}) - \hat{e})]}_{\text{Effect on behavior}} + \underbrace{\alpha(e^*(\hat{e}) - \hat{e})}_{\text{Effect on preferences}}$$

Consider the expression labelled 'Effect on behavior.' Plugging in the optimal action e^* derived in expression 3 yields:

$$\Delta - ce^*(\hat{e}) - \alpha(e^*(\hat{e}) - \hat{e}) = \Delta - c\left(\frac{\Delta + \alpha\hat{e}}{\alpha + c}\right) - \alpha\left(\frac{\Delta + \alpha\hat{e}}{\alpha + c} - \hat{e}\right) = 0.$$

The result follows because a marginal change in the identity norm $\hat{e}$ is fully internalized: effort changes in such a way that it exactly 'offsets' the change in the identity norm. By contrast, the direct effect on preferences of an increase in identity norm $\hat{e}$ can be positive or negative, depending on parameter values. The following result describes the sign of the effect on preferences:

Lemma 1. *The marginal effect of a more demanding identity norm on the member's equilibrium utility is positive if* $\hat{e} < \frac{\Delta}{c}$*, negative if* $\hat{e} > \frac{\Delta}{c}$*, and equal to* 0 *if* $\hat{e} = \frac{\Delta}{c}$.

Several observations follow from this result. First, there is only scope for identity change that moves the identity norm upward if the pre-existing norm is relatively less demanding; that is, $\hat{e}_E < \frac{\Delta}{c}$. Otherwise, the citizen will demand norms that are less demanding than the pre-existing norm. Second, the norm $\frac{\Delta}{c}$ maximizes the equilibrium utility of the citizen, and his combined material and psychological utility is symmetrically decreasing around this point.[13] This is formally stated in the next result, which shows the connection between material and identity concerns:

[13]More precisely, the function V is a polynomial of degree 2 with respect to $\hat{e}$. Consequently, it is first increasing and then decreasing in $\hat{e}$.

Corollary 1. *The citizen's optimal identity norm is equal to the optimal choice when there are* no *identity concerns, that is:*

$$\hat{e}_{\text{opt}} = e^*(\alpha = 0) = \frac{\Delta}{c}$$

As a consequence:

$$e^*(\hat{e}_{\text{opt}}) = \frac{\Delta + \alpha(\frac{\Delta}{c})}{\alpha + c} = \frac{\Delta}{c}$$

Proof. Follows directly from Lemma 1.

The intuition for these results is as follows. The citizen's total utility function U consists of a combination of material (U^M) and identity concerns (U^I). For some norms, there might be tension between those, requiring effort to be an uneasy compromise between material and identity interests. The optimal norm is that such tension is eliminated so that the resulting effort choice maximizes both material and identity interests.

Instead of looking at an arbitrary identity norm, now consider the pre-existing norm $\hat{e}_E$ and the proposed norm $\hat{e}_P$. When considering whether or not to affirm, the citizen considers the equilibrium utility of each norm and affirms if, and only if, the equilibrium utility of the proposed norm is weakly higher than the equilibrium utility of the existing norm; that is:

$$V(\hat{e}_P) \geq V(\hat{e}_E).$$

The following result outlines the conditions under which this inequality is satisfied.[14]

Lemma 2. *If* $\hat{e}_P \geq \hat{e}_E$, *then the group member affirms* $\hat{e}_P$ *if*

$$\hat{e}_P \leq \frac{2\Delta}{c} - \hat{e}_E$$

The reverse inequality determines if $\hat{e}_P$ *is affirmed when* $\hat{e}_P \leq \hat{e}_E$.

Lemma 2 describes the citizen's affirmation constraint. It says that if a pre-existing identity norm is sufficiently low ($\hat{e}_E < \hat{e}_{\text{opt}}$), then the citizen affirms new identity norms whenever it is closer to the optimal identity norm. This is the case if the norm is sufficiently low ($\hat{e}_P \leq 2\frac{\Delta}{c} - \hat{e}_E$).

Now consider the identity norm the propagandist wishes to propose, given the range of acceptable identities. Since the propagandist is only interested in obtaining the outcome $y = 1$, she wishes to instill the highest level of effort possible. By expression 3, the effort is higher when the identity norm is more demanding. This means that the propagandist propagates the most demanding identity possible that is still affirmed; that is, $\hat{e}_P^* = 2\frac{\Delta}{c} - \hat{e}_E$ if there is scope for identity change that results in more demanding norms ($\hat{e}_E < \hat{e}_{\text{opt}}$). If the pre-existing norm is already relatively demanding ($\hat{e}_E > \hat{e}_{\text{opt}}$), more demanding norms will not be affirmed. Then the propagandist proposes either the pre-existing norm or an identity norm that will not be affirmed. Summarizing this discussion:

[14]Rearranging, the inequality could also be written as $V(\hat{e}_P) - V(\hat{e}_E) \geq 0$. Because the citizen considers the difference in equilibrium utilities, parameters that do not interact with effort or the identity norm do not affect the affirmation choice. In particular, adding a positive constant to the function U^I does not affect the analysis.

Proposition 1. *In the equilibrium of the game, if existing identity content is relatively less demanding* ($\hat{e}_E \leq \hat{e}_{\text{opt}}$), *the propagandist proposes* $\hat{e}_P^* = 2\frac{\Delta}{c} - \hat{e}_E$, *which is affirmed by the citizen. Optimal propaganda becomes more demanding as the citizen's stakes increase* $\left(\frac{\partial \hat{e}_P^*}{\partial \Delta} > 0\right)$, *less demanding as mobilization costs increase* $\left(\frac{\partial \hat{e}_P^*}{\partial c} < 0\right)$, *and less demanding as the pre-existing identity norm increase* $\left(\frac{\partial \hat{e}_P^*}{\partial \hat{e}_E} < 0\right)$.

The analysis so far is illustrated in Fig. 1, which displays, for two different levels of mobilization costs, c, the range of identity norms that the citizen affirms, the optimal identity norm, and the equilibrium identity norm.

Before proceeding, it is worth discussing two variations of the baseline model. First, consider the case in which the citizen's material interests are strictly the opposite of the leaders; that is, the citizen is better off when the outcome is $y = 0$ instead of $y = 1$ (so that $u(0) > u(0)$ and $\Delta < 0$). The analysis goes through as before with $e^* = \frac{\Delta + \alpha\hat{e}}{\alpha + c}$ or 0 if Δ is smaller than $\alpha\hat{e}$. The derivative of the function $V(\hat{e})$ is positive if $\frac{\Delta}{c} > \hat{e}$, as before. However, this inequality can never be satisfied because the left-hand side is negative. Thus, the citizen will never affirm more demanding identity norms. Hence, the partial alignment of propagandists' and citizens' material preferences is an important scope condition. I show below that this requirement can be relaxed when the propagandist can allocate additional material resources.

Second, it is straightforward to generalize the analysis to include more than one citizen as long as the citizens are identical and there is no interaction of effort levels at the mobilization stage. Specifically, in the Appendix, I solve for the equilibrium of the game in which there are n citizens, indexed by i, and the technology is given by $\Pr(y = 1|e_1, \ldots, e_n) = \frac{\sum_{i=1}^{n} e_i}{n}$. The analysis goes through as before, slightly adjusting all terms to take into account the reduced effectiveness of effort, which is now $\frac{1}{n} < 1$.

Empirical Implications

The model can provide insights into a range of phenomena. In this subsection, I discuss several examples and how the model can explain empirical findings that the quantitative literature on propaganda has accumulated.[15]

Covid-19: An important application concerns the emergence of COVID-19 in early 2020. Following the initial outbreak, some leaders, especially those in China, urged citizens to change their habits, to start social distancing, wear masks, and adhere to lockdowns. In terms of the model, these behaviors could be interpreted as exerting high effort to avoid a large-scale outbreak (where an outbreak is the outcome $y = 0$), which many citizens plausibly preferred; hence, the assumption that $u(1) > u(0)$ is satisfied here. However, existing norms ($\hat{e}_E$) were mostly inconsistent with these novel public health measures. Some leaders tried to tie new norms to identities, which can be interpreted as proposing a high level of $\hat{e}_P$. Empirical research finds patterns consistent with the model. In particular, norm-based masking messages can be effective (Bokemper et al. 2022; Raymond, Kelly, and Hennes 2021) and are more effective for citizens that plausibly have a higher stake in the outcome (Grossman et al. 2020). Interestingly, the model points out that relatively less demanding existing norms ($\hat{e}_E$ small) are a necessary condition for effective communication. Thus, the fact that existing norms were relatively inconsistent with this new public health guidance may have enabled (more) effective leader communication.

Inter-group conflict: Another implication of the analysis concerns the social construction of identities during times of conflict (Fearon and Laitin 2000). Scholars have extensively

[15]It is important to note that the analysis above relied on a symmetric loss function; doing too much is as costly as doing too little. This is plausible for some, but not all, empirical applications. See the conclusion section for further discussion.

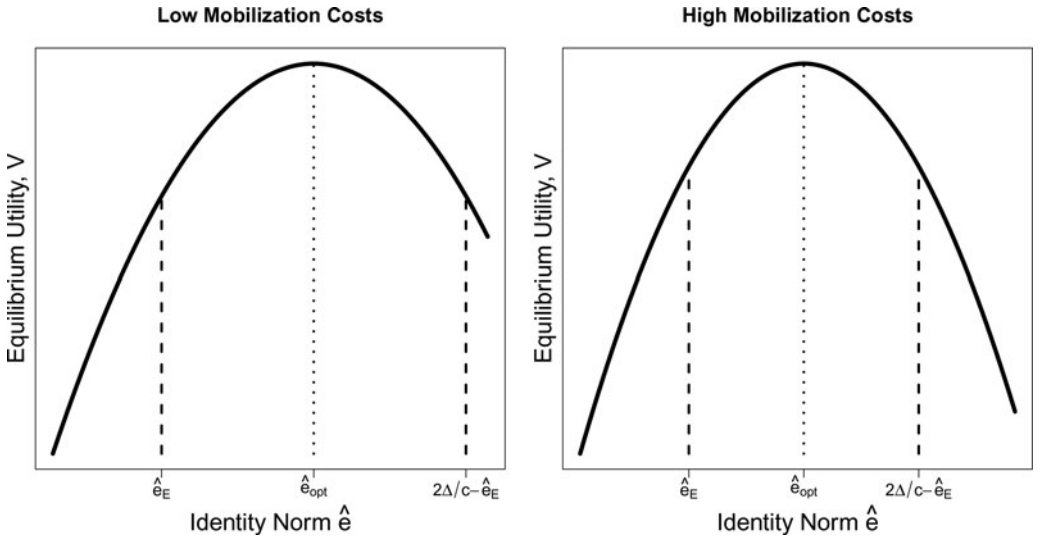

Figure 1. The citizen's equilibrium utility as a function of different identity norms for low mobilization costs (left panel) and high mobilization costs (right panel). Parameter values: $\Delta = 1$, $\alpha = 0.3$,, and $\hat{e}_E = 0.25$. Left panel: $c = 1.5$. Right panel: $c = 2$.

documented that elites can often manipulate citizens' identities in their favor (for example, Brass 1997), the outcome of which is that citizens take costly actions that ultimately keep elites in power. For Fearon and Laitin (2000), this pattern constitutes a puzzle because citizens seem to be willing to follow elites down paths that primarily benefit elites. Building on their review of related work, they conjecture that citizens follow elites because they have their own agendas, which may have 'little to do with communal antipathies per se' (Fearon and Laitin 2000, 855). In terms of my model, effective identity propaganda requires that citizens' material incentives are aligned with the leader; that is, $u(1) > u(0)$. This can be interpreted as the citizen having an 'agenda' but the terms need not correspond to any kind of ethnicity-based motivations. For example, Woodward (1995) emphasizes that the inter-group conflict in Bosnia became a rare opportunity for enrichment in a period of serious economic hardship.

Furthermore, in the equilibrium analyzed here, the propagandist's agenda-setting power allows her to propose identity prescriptions that are too demanding relative to the citizen's most-preferred identity norm. The equilibrium proposal keeps the citizen exactly indifferent between the existing identity norm and the new norm while motivating substantially higher equilibrium effort by the citizen – and generating a substantially higher payoff for the propagandist. Thus, while citizens 'follow,' that is, exert high effort due to propaganda, elites leverage their strategic position to manipulate the citizen's behavior in their favor.

Finally, the model also emphasizes the mobilization costs, c, as an important determinant of propaganda; the lower it is, the more demanding the (affirmed) equilibrium identity norm. Consistent with this result, the case studies in Fearon and Laitin (2000) mention the availability of young men, presumably endowed with low mobilization costs, for the importance of effective identity propaganda and, ultimately, conflict. For example, Kapferer (2011) mentions that the prevalence of gangs – whose members were often impoverished and unemployed youth – is an important determinant of propaganda and conflict. In sum, my model provides a micro-foundation for the conjectures and empirical findings mentioned in Fearon and Laitin (2000).

National crises and the rally-'round-the-flag effect: The model can also shed light on instances of national crises and rapidly rising incumbent support, which is known as the 'rally-'round-the-flag' phenomenon (Baker and Oneal 2001). In terms of the model, the effort term e could be interpreted as any costly action that supports the incumbent. Many studies

interested in understanding the rally-'round-the-flag phenomenon manipulate the severity of the threat (Kobayashi and Katagiri 2018), which can be understood as changing the stakes parameter Δ. The model shows that the proclivity to affirm more demanding identity norms increases as this parameter increases. Anticipating this effect, the incumbent airs more demanding identity norms, which increases efforts. Thus, the model explains the finding documented in the literature that the severity of threat perceptions can cause more extreme rhetoric by incumbents and, eventually, support for incumbents.

Quantitative literature on propaganda: The baseline model demonstrates that under some conditions – for example, if the citizen has a positive stake in the outcome and existing identity content is relatively less demanding – identity propaganda effectively influences a citizen's effort choice. This is consistent with the recent empirical literature that uses identification strategies to demonstrate that exposure to propaganda has a causal effect on citizens' behavior (Adena et al. 2015; Barber and Miller 2019; Bleck and Michelitch 2017; Blouin and Mukand 2019; DellaVigna et al. 2014; Enikolopov, Petrova, and Zhuravskaya 2011; Mitts, Phillips, and Walter 2022; Peisakhin and Rozenas 2018; Yanagizawa-Drott 2014). As discussed above, this literature takes the design of propaganda as given and studies the consequences of being exposed to it. However, Proposition 1 reveals that leaders design propaganda differently depending on the context. It is more demanding the higher the citizen's stakes in the outcome (Δ) are, but less demanding when mobilization costs are high (c), or when existing norms are already relatively demanding ($\hat{e}_E$). Thus, while all studies are concerned with identifying the causal effect of propaganda, and identity messages are presumably part of every broadcast, the exact content varies across studies. Hence, it is difficult to compare their persuasive effects. In the following sections, I continue to probe the relationship between propaganda design and important features of the environment. I consider preference heterogeneity, strategic interaction, and the presence of material incentives in turn.

Applications

Heterogeneity of Citizens' Preferences

In the real world, propagandists often face a heterogeneous audience. In this section, I incorporate this fact by introducing a second citizen who is also a member of the same social group, but whose stakes regarding the outcome differ. I demonstrate that for the propagandist, heterogeneity across citizens induces a novel tradeoff between proposing moderate, universally affirmed identity norms and more demanding, partially affirmed ones, i.e., 'divisive norms.' Perhaps surprisingly, I find that the latter option is optimal for the propagandist if existing norms are sufficiently demanding and mobilization costs are high, which can potentially explain the behavior of populist politicians attempting to mobilize relatively marginalized groups (Anduiza, Guinjoan, and Rico 2019; de la Torre 2017). However, since both citizens are less likely to affirm demanding identity norms under these conditions, the overall effect of either an increase in mobilization costs or more demanding pre-existing norms on equilibrium identity norms is non-monotone.

I denote the two citizens by a and b, and refer to an arbitrary citizen by i. Each citizen can exert effort $e_i \in [0, 1]$ to bring about the desired political outcome. Their effort choices are pure substitutes, as in a public good game. In particular, the probability that the outcome $y = 1$ occurs is:

$$\Pr(y = 1|e_a, e_b) \equiv \gamma(e_a, e_b) = \frac{e_a + e_b}{2}. \tag{4}$$

Turning to the equilibrium at the mobilization stage, each group member $i \in \{a, b\}$ solves the following maximization problem:

$$\max_{e_i \in [0,1]} \gamma\Delta_i - \frac{c}{2}e_i^2 - \frac{\alpha}{2}(e_i - \hat{e}_i)^2$$

I assume that initially, the two group members hold identical identity norms: $\hat{e}_{aE} = \hat{e}_{bE} = \hat{e}_E$. Thus, heterogeneity across group members is limited to the stakes, Δ_i. Without loss of generality, I assume that group member a is the one with the higher stakes; that is, $\Delta_a > \Delta_b$. I interpret this to mean that citizen a is an 'extremist,' and citizen b a 'moderate.' Finally, I assume pre-existing content is relatively less demanding: $\hat{e}_E < 2\frac{\Delta_b}{c}$, which implies that there is scope for identity change for both citizens.[16]

Similar to the above, each group member's optimal effort choice is given by:

$$e_i^*(\hat{e}_i) = \frac{\frac{\Delta_i}{2} + \alpha\hat{e}_i}{\alpha + c}.$$

Again, each citizen's level of effort is higher when his internalized identity norm is more demanding.

If citizen i affirms new identity content, his effort changes while the other citizen's effort remains the same (i's effort is independent of the identity norm of citizen $-i$). The expected utility of affirming the content proposed by the propagandist, $\hat{e}_{iP}$, is:

$$\frac{e_i^*(\hat{e}_{iP}) + e_{-i}^*}{2}\Delta_i - \frac{c}{2}(e_i^*(\hat{e}_{iP}))^2 - \frac{\alpha}{2}(e_i^*(\hat{e}_{iP}) - \hat{e}_{iP})^2 \equiv V_i(\hat{e}_{iP})$$

While the expected utility of not affirming new content is:

$$\frac{e_i^*(\hat{e}_E) + e_{-i}^*}{2}\Delta_i - \frac{c}{2}(e_i^*(\hat{e}_E))^2 - \frac{\alpha}{2}(e_i^*(\hat{e}_E) - \hat{e}_E)^2 \equiv V_i(\hat{e}_E)$$

Examining the inequality $V_i(\hat{e}_{iP}) \geq V_i(\hat{e}_E)$ yields the following result:

Lemma 3. *If $\hat{e}_{iP} \geq \hat{e}_E$, citizen i accepts a new identity norm if*

$$\hat{e}_{iP} \leq \frac{\Delta_i}{c} - \hat{e}_E$$

If $\hat{e}_{iP} \leq \hat{e}_E$, the reverse inequality describes the affirmation constraint.

Note that citizen i's affirmation rule is independent of the choice of the other group member and qualitatively unchanged from the baseline case. Moreover, given that the affirmation constraint is a function of each citizen's stakes, it is 'looser' for citizen a than for citizen b. In other words, citizen a is more willing to affirm more demanding norms than citizen b. This is illustrated in Fig. 2, left panel.

Given citizens' differing affirmation constraints and the fact that citizen a's stakes are larger, $\Delta_a > \Delta_b$, the propagandist faces a tradeoff. She can propose $\frac{\Delta_b}{c} - \hat{e}_E \equiv \hat{e}_*$, in which case both group members will accept the new identity norm. Alternatively, she can propose $\frac{\Delta_a}{c} - \hat{e}_E \equiv \hat{e}_{**}$, in which case only citizen a will accept the new norm: citizen b will view the norm as unacceptable and keep his existing identity content $\hat{e}_E$. The propagandist chooses the more demanding (or 'divisive') content $\hat{e}_{**}$ if the probability of obtaining the outcome $y = 1$ given this norm is larger than the corresponding probability given the more moderate norm $\hat{e}_*$, that is:

$$\frac{e_a^*(\hat{e}_{**}) + e_b^*(\hat{e}_E)}{2} \geq \frac{e_a^*(\hat{e}_*) + e_b^*(\hat{e}_*)}{2}.$$

[16] As mentioned above, when there is more than one citizen, the condition for effective communication has to be adapted to consider the effort's reduced effectiveness.

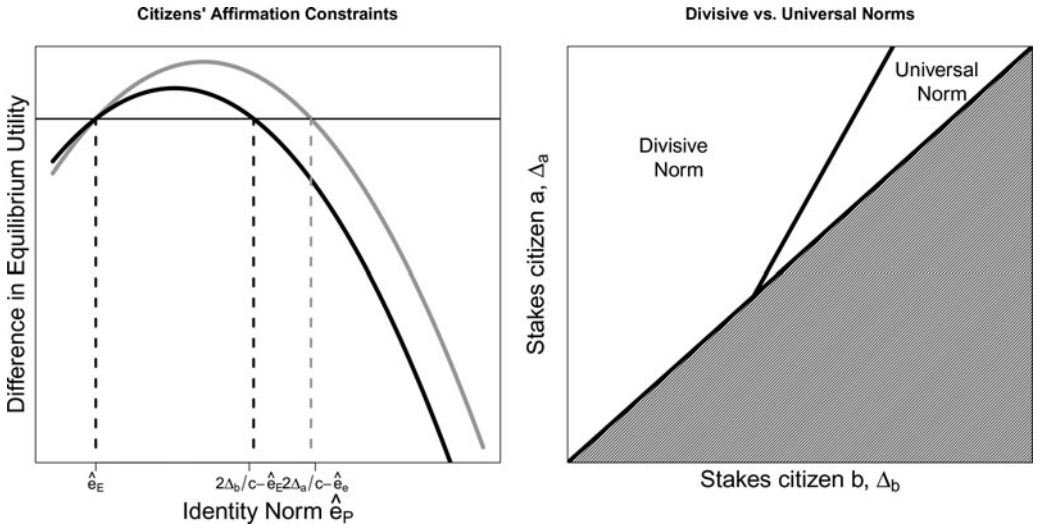

Figure 2. Left panel: Difference in equilibrium utility $V_i(\hat{e}_P) - V_i(\hat{e}_E)$ for a moderate ($i = b$; black line) and an extremist group member ($i = a$, gray line). Parameter values: $\Delta_a = 1.05$, $\Delta_b = 0.85$, $c = 1.5$, $\alpha = 0.3$, and $\hat{e}_E = 0.1$. Right panel: decision rule of the propagandist as a function of group members stakes Δ_a and Δ_b. Parameter values: $c = 1.5$ and $\hat{e} = 0.2$. The area shaded in gray is ruled out by the assumption that $\Delta_a > \Delta_b$.

Plugging in the relevant quantities and re-arranging yields:

$$\Delta_a \geq 2[\Delta_b - c\hat{e}_E] \tag{5}$$

This decision rule is illustrated in Fig. 2, right panel. Examining the preceding inequality yields the following result:

Proposition 2. *An increase in mobilization costs c or in the pre-existing identity norm $\hat{e}_E$ increases the likelihood that the divisive norm will be aired.*

To see the intuition of this result, consider a more demanding existing norm first. When the propagandist chooses the divisive norm, citizen a's norm is $\hat{e}_{**} = \frac{\Delta_a}{c} - \hat{e}_E$ and citizen b's norm is $\hat{e}_E$. As a result of the linearity (both of the technology γ and the optimal effort e_i^* choices with respect to i's norm), the existing norm cancels out. By contrast, when the propagandist chooses the universal norm, $\hat{e}_* = \frac{\Delta_b}{c} - \hat{e}_E$, a more demanding existing norm decreases the newly proposed norm, which decreases effort. Hence, airing the divisive norm $\hat{e}_{**}$ is optimal. Substantively, a more demanding existing norm makes it less bad that citizen b does not affirm but, instead, simply chooses effort in accordance with the existing norm.

Now consider an increase in mobilization cost c. This negatively affects both candidate's choices, $\hat{e}_*$ and $\hat{e}_{**}$, but does not affect citizen b's identity norm when the divisive norm is aired (because b rejects $\hat{e}_{**}$ and keeps $\hat{e}_E$). Moreover, due to the functional forms employed, mobilization costs, c, affect both candidate norms in a similar fashion, which ultimately means that the effect is stronger on the moderate, universally-affirmed norm $\hat{e}_*$. As a consequence, the propagandist is better off choosing the divisive, partially-affirmed norm $\hat{e}_{**}$.[17]

[17]It is clear that linearity plays an important part in this result – both in terms of the technology and how norms matter for optimal effort. An alternative specification for γ would feature (strict) concavity. This could affect the propagandist's incentives to air the divisive norm. In particular, choosing the divisive norm could have a smaller effect on the probability of

The literature on populism has found some support for the prediction documented in Proposition 2. In particular, populist leaders in a range of polities – Venezuela, Turkey, and the US – have mobilized previously marginalized voters; that is, faced relatively high mobilization costs using divisive rhetoric (Anduiza, Guinjoan, and Rico 2019; Cheeseman and Larmer 2015; de la Torre 2017; McCoy and Somer 2019).[18] For example, former US President, Donald Trump, implored his supporters to 'fight like hell' even in the face of increasingly high costs when trying to overturn the results of the 2020 presidential election (NPR, 2021). Research has shown that such divisive rhetoric can substantially change political outcomes (Newman et al. 2021).

In general, however, Lemma 3 and Proposition 2 imply that mobilization costs (and pre-existing norms) have a non-monotone effect on equilibrium norms. This is illustrated in Fig. 3.

The intuition is as follows. As can be seen from Lemma 3, an increase in either factor decreases the norms that citizens find acceptable; everything as being equal, this effect decreases the equilibrium norm. However, after a certain point (as described in expression 5), the propagandist strategically refrains from airing universal norms and switches to divisive norms, rendering equilibrium norms more demanding. Both mobilization costs (and existing identity norms) are difficult to measure. However, even assuming that they could be accurately measured, as a consequence of this non-monotone relationship, future empirical scholarship attempting to measure the effect of mobilization costs on identity norms (or pre-existing identity norms) may find a positive or negative effect – depending on the cost levels that form the bases of the comparisons.

I conclude this section with a brief discussion on how one can use the model to investigate the effect of the sizes of different factions. In the Appendix, I generalize the analysis so that there are n citizens, divided into factions of potentially unequal sizes. Specifically, out of the n citizens, k are assumed to be members of the faction a whereas $n-k$ citizens are members of the more moderate faction b. The propagandist still decides between the moderate, universally affirmed norm and the more demanding, divisive norm. The generalization of expression 5 is:

$$\Delta_a \geq \frac{n}{k}[\Delta_b - (n-k)\hat{e}_E c].$$

Clearly, the left-hand side is decreasing in k: the higher the number of extremists, the more attractive the divisive norm $\hat{e}_{**}$ becomes. However, depending on parameter values, choosing the divisive norm can be optimal even though moderate material interests characterize most group members. In particular, I show in the Appendix that when the difference in stakes $\Delta_a - \Delta_b$ is large, the divisive norm can emerge in equilibrium even though the extremists are in the minority. Thus, a relatively small, extreme minority can drive the emergence of harsh identity norms and behavior.

Strategic Interaction among Citizens

Existing empirical work suggests that strategic interaction among citizens – and especially coordination incentives – is often crucial for upholding norms (Invernizzi et al. 2021). In this section, I adapt the mobilization stage to include richer strategic incentives for citizens. I show that strategic interaction at the mobilization stage induces multiple equilibria at the affirmation stage. This has two important implications. First, it complicates the propagandist's calculus, having to take into account not only each citizen's affirmation constraint but also their expectations about other citizen's affirmation constraints. Second, multiple equilibria can explain why actors, organizations, or polities that seem similar in terms of observable attributes can end up with very different identity

obtaining the outcome $y = 1$, and hence on the propagandist's utility. Everything else being equal, this should decrease the attractiveness of pushing for the more demanding, divisive identity norm.

[18]That this rhetorical strategy is not inherent to populism is exemplified by research on populist politicians in Asia (Pepinsky 2020). Because of the different political environments, populists do not choose divisive rhetoric.

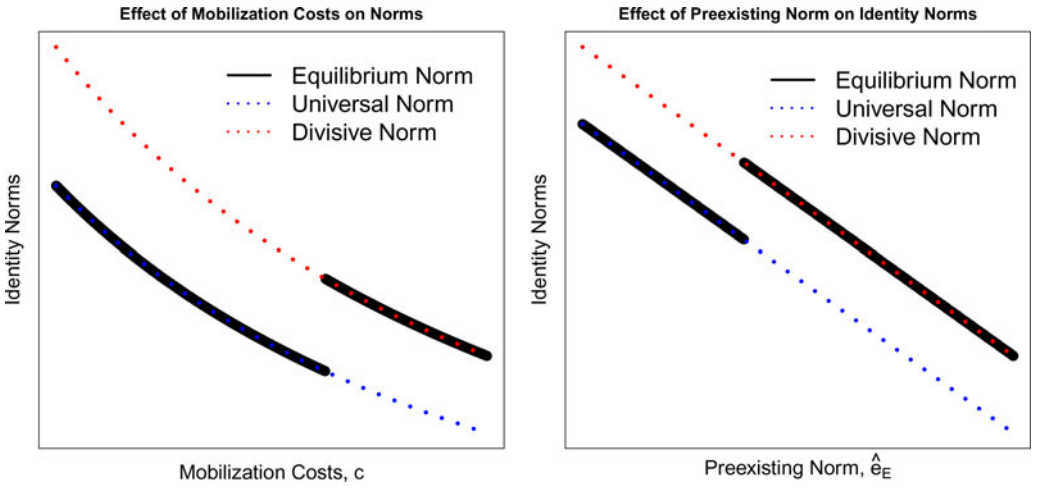

Figure 3. Effect of Mobilization Costs (left panel) and pre-existing norms (right panel) on identity norms. Parameter values: $\Delta_a = 1.25$, $\Delta_b = 1$. Left panel: $\hat{e}_E = 0.2$ and $c \in [\Delta_b, \Delta_b + 1]$. Right panel: $c = 1.3$, and $\hat{e}_E \in [0, (\Delta_b/c)]$.

norms (Carrillo and Gromb 1999; Gibbons 2010) and, ultimately, success rates (Amenta et al. 2010).

As before, there are two citizens, indexed by $i = a, b$. In contrast to the model variation analyzed in the previous subsection, I assume here that the citizens' stakes are equal, that is, $\Delta_a = \Delta_b \equiv \Delta$. However, the technology is now given by:

$$\Pr(y = 1|e_a, e_b) \equiv \gamma(e_a, e_b) = \gamma_n(e_a + e_b) + \gamma_s e_a e_b.$$

The term $\gamma_n > 0$ scales non-strategic incentives whereas the term γ_s represents strategic incentives. If $\gamma_s > 0$, the mobilization stage features strategic complements, whereas if $\gamma_s < 0$, the game features strategic substitutes. I assume that $\gamma_n > |\gamma_s|$ to ensure citizens choose positive levels of effort in equilibrium. As before, I assume that the group members choose to affirm or not simultaneously. Let $t = (t_a, t_b)$ be the vector of affirmation choices.

Consider first the equilibrium at the effort stage. Each group member i solves:

$$\max_{e_i \in [0,1]} \gamma(e_i, e_{-i})\Delta - \frac{c}{2}e_i^2 - \frac{\alpha}{2}(e_i - \hat{e}_i)^2$$

This maximization problem implies the following best response function:

$$e_i(e_{-i}) = \frac{\gamma_n + \gamma_s e_{-i}}{\alpha + c}\Delta + \frac{\alpha}{\alpha + c}\hat{e}_i.$$

Note that citizen i's best response is increasing in e_{-i} if $\gamma_s > 0$ and decreasing in e_{-i} if $\gamma_s < 0$, as expected. Moreover, i's best response is a function of i's identity norm only.

Solving the pair of best response functions yields the equilibrium effort levels:

$$e_i^* = E_0 + E_1\hat{e}_i + E_2\hat{e}_{-i}, \tag{6}$$

where the constants E_0, E_1, and E_2 are defined in the Appendix. Intuitively, citizen i's equilibrium effort is a linear function of both i's and $-i$'s (actual, affirmed) identity norm. The sign of E_2 is

equal to the sign of γ_s, so that when citizen $-i$'s identity norm becomes more demanding, i's effort increases (decreases) if the game features strategic complements (substitutes).

The next step is to solve for the equilibrium at the affirmation stage. It is again useful to consider the value of i's identity norm, holding $-i$'s identity norm fixed:

$$V_i(\hat{e}_i, \hat{e}_{-i}) = \gamma(e_i^*, e_{-i}^*)\Delta - \frac{c}{2}(e_i^*)^2 - \frac{\alpha}{2}(e_i^* - \hat{e}_i)^2$$

where all equilibrium actions are a function of both i's and $-i$'s respective identity norms, as outlined in expression 6. Consider the derivative of V_i with respect to i's identity norm:

$$\begin{aligned}\frac{\partial V_i}{\partial \hat{e}_i} &= \left(\frac{\partial \gamma}{\partial e_i}\frac{\partial e_i^*}{\partial \hat{e}_i} + \frac{\partial \gamma}{\partial e_{-i}}\frac{\partial e_{-i}^*}{\partial \hat{e}_i}\right)\Delta - ce_i^*\frac{\partial e_i^*}{\partial \hat{e}_i} - \alpha(e_i^* - \hat{e}_i)\left(\frac{\partial e_i^*}{\partial \hat{e}_i} - 1\right) \\ &= \underbrace{\frac{\partial e_i^*}{\partial \hat{e}_i}\left[\frac{\partial \gamma}{\partial e_i}\Delta - ce_i^* - \alpha(e_i^* - \hat{e}_i)\right]}_{\text{Effect on own behavior}} + \underbrace{\frac{\partial \gamma}{\partial e_{-i}}\frac{\partial e_{-i}^*}{\partial \hat{e}_i}\Delta}_{\text{Effect on other behavior}} + \underbrace{\alpha(e_i^* - \hat{e}_i)}_{\text{Effect on preferences}}\end{aligned}$$

There are three effects detailed in the preceding expression. As before, when looking at the marginal effect of a more demanding identity norm, the 'Effect on own behavior' is 0 while the effect on preferences can be positive or negative. In addition, there is now a third effect: affirming a more demanding identity norm has positive or negative implications for $-i$'s behavior because depending on the sign of γ_s, a more demanding identity norm can increase or decrease the other citizen's effort level. This effect stems from the presence of richer strategic incentives at the mobilization stage.

I now investigate what kind of identity content is accepted; that is, under which conditions does the inequality $V_i(\hat{e}_{iP}, \hat{e}_{-i}) \geq V_i(\hat{e}_{iE}, \hat{e}_{-i})$ hold. Similar to the baseline model, the inequality can be re-arranged to reveal that citizen i accepts an interval of identity norms. Specifically:

Lemma 4. *Suppose that some identity content is affirmed* ($\frac{\partial V_i}{\partial \hat{e}_i}|_{\hat{e}_i=\hat{e}_{-i}=\hat{e}_E} > 0$) *and that* $\hat{e}_P > \hat{e}_E$. *There exist constants* T_0 *and* T_1 *such that* i *affirms* $\hat{e}_P$ *if and only if:*

$$\hat{e}_P \leq T_0 + T_1 \cdot \hat{e}_{-i}(t_{-i}),$$

where $\hat{e}_{-i}(t_{-i})$ *is the norm by the other citizen,* $-i$, *that depends on the conjectured affirmation choice by* $-i$, t_{-i}. *Importantly, the sign of* T_1 *is equal to the sign of* γ_s.

To see the implications of Lemma 4, consider the following two cases. Consider the case of strategic complements first, that is, $\gamma_s > 0$, which implies $T_1 > 0$. Substantively, this means that citizen i is willing to accept more demanding identity content if i expects $-i$ to do the same. This means that the range of affirmed identity content is wider if $-i$ also affirms the new identity content. Now consider the case of strategic substitutes, that is, $\gamma_s < 0$, which implies $T_1 < 0$. Now citizen i is less willing to accept more demanding identity if i expects $-i$ to affirm. In other words, the range of affirmed content is tighter if $-i$ affirms new identity content.

The analysis so far reveals that the mobilization stage has spillover effects on the affirmation stage. To investigate this further, define $\hat{e}' \equiv T_0 + T_1\hat{e}_E$ and $\hat{e}''$ as the solution to

$$\hat{e}'' = T_0 + T_1\hat{e}'' \;\Rightarrow \hat{e}'' \equiv \frac{T_0}{1 - T_1}.$$

The interpretation of these two values depends on the nature of strategic interaction. Again, consider the case of strategic complements first. Here, if $\hat{e}' > \hat{e}_E$, then $\hat{e}'' > \hat{e}'$ and consider the

following cases.[19] First, suppose that the proposed $\hat{e}_P$ is smaller than $\hat{e}'$. This means that each i will affirm even if they expect the other group member not to affirm. As a result, the affirmation stage is dominance solvable and each player chooses to affirm: $t^\star = (1, 1)$. Second, suppose that $\hat{e}_P > \hat{e}''$. Now, neither player affirms, even if they expect the other player to affirm: $t^\star = (0, 0)$. Third, suppose that the proposed $\hat{e}_P$ is in an intermediate range; that is, in between $\hat{e}'$ and $\hat{e}''$. In this case, there are multiple (pure strategy) equilibria because expectations can be self-fulfilling. If i expects $-i$ not to affirm, neither will i and so $t^\star = (0, 0)$. However, if i expects $-i$ to affirm, so will i and so $t^\star = (1, 1)$.

Now consider the case of strategic substitutes. Here, if $\hat{e}' > \hat{e}_E$, then $\hat{e} > \hat{e}''$ and consider the following cases. First, suppose that the proposed $\hat{e}_P$ is smaller than $\hat{e}''$. The affirmation stage is again dominance solvable and both group member affirm, so that $t = (1, 1)$ is the unique equilibrium. Moreover, if $\hat{e}_P > \hat{e}'$, then both citizens choose not to affirm. In the intermediate range in which $\hat{e}_P$ is between $\hat{e}''$ and $\hat{e}'$, there are multiple asymmetric equilibria at the affirmation stage, so that $t = (1, 0)$ and $t = (0, 1)$ are both equilibria. The reason is that if citizen i expects citizen $-i$ to affirm, they do not wish to affirm themselves. Conversely, citizen $-i$ expects i not to affirm, so they are willing to affirm themselves. In this range, citizens become endogenously differentiated, even though they are ex ante identical; that is, have the same stakes, mobilization costs, and salience of identity.

This latter finding has interesting substantive implications when the model is applied to revolutionary politics. In contrast to the case of strategic complements, if they are multiple equilibria, they differ not in terms of total effort exerted but, rather, in who exerts a high effort and who exerts a low effort. Due to self-fulfilling expectations, one citizen becomes a 'vanguard,' exerting high effort, while the other becomes a 'follower,' exerting lower levels of effort. Most theories of vanguards assume that vanguards display some special characteristic that makes them more skilled revolutionaries (e.g., Bueno De Mesquita 2010).[20] Here, however, citizens are ex ante identical and only become differentiated – in terms of identity norms and equilibrium effort levels – due to the strategic interaction inherent in the game.

Finally, consider the propagandist's choice. The propagandist wishes to maximize the effort levels exerted. Differentiating the propagandist's objective function, it is the case that independent of the sign of γ_s, the propagandist chooses either $\hat{e}'$ or $\hat{e}''$; that is, the endpoints of the intervals of affirmed identity norms. When effort choices are strategic complements, the coordination problem at the affirmation stage looms especially large. If the group members are uncoordinated, that is, $t^\star = (0, 0)$ for any $\hat{e}_P > \hat{e}'$, then the best the propagandist can do is to propose $\hat{e}_P^* = \hat{e}'$. If the group members are coordinated so that $t^\star = (1, 1)$ even for more demanding identity norms, the propagandist proposes $\hat{e}_P^* = \hat{e}''$. Summarizing:

Proposition 3. *Suppose that effort levels are strategic complements. There are two pure strategy equilibria. Equilibrium identity content will be more demanding if the group members are* coordinated *at the affirmation stage.*

When effort levels are strategic substitutes, the propagandist faces a similar tradeoff to the case in which there are heterogeneous preferences. In particular, the leader chooses between inducing a moderate norm that is affirmed by all and a more demanding norm that is affirmed by one group member only. However, this is not driven by intrinsic differences (in terms of preferences or information) between citizens, but rather by the strategic incentives among symmetrically situated citizens. I show in the Appendix that for a wide range of parameter values, the propagandist choose the former option, airing the moderate, universally accepted norm.

[19]For proof of this claim, see the Appendix.

[20]My notion is related to the 'early risers' definition of a vanguard (for a comparison of different vanguard definitions, see Shadmehr and Bernhardt 2019).

In general, the analysis demonstrates that when there are rich strategic incentives at the mobilization stage, they have spillover effects for affirming proposed identity norms. Thus, otherwise identical actors, organizations, or polities can exhibit very different identity norms. From research in sociology, economics, and political science, there are two pieces of evidence that speak to the results. First, research across these fields demonstrates that seemingly similar organizations can end up with very different norms or, more generally, 'cultures' (Carrillo and Gromb 1999; Gibbons 2010). For example, scholars have shown that organizations as disparate as firms and social movements develop distinct behavioral codes (Akerlof and Kranton 2000). Second, research also shows that seemingly similar organizations can have a very different 'effectiveness,' both in terms of intermediate and ultimate policy goals (Amenta et al. 2010). For example, there is significant variation in rebel organization to avoid civilian casualties, even among groups with similar observable attributes (Gibilisco, Kenkel, and Rueda 2022; Humphreys and Weinstein 2006).

Interaction with Material Incentives

In many situations, leaders attempt to shape citizen action by broadcasting propaganda and by providing direct material incentives in the form of repression ('sticks') or cooptation ('carrots') (Gehlbach, Sonin, and Svolik 2016). In this section, I consider an expanded model in which the propagandist can simultaneously propose new identity norms and use material incentives, denoted by $r \in [0, \bar{r}]$, to motivate citizen effort. Choosing a higher level of material incentives is more costly, so her cost function $\Psi(r)$ is increasing, convex, and satisfies standard Inada conditions.[21] I show that a higher level of material incentives, r, relaxes the citizen's affirmation constraint and enables the propagandist to choose more demanding identity norms – which also enhances the effectiveness of material incentives. Thus, the citizen's affirmation calculus is a source of complementarity between material incentives and propaganda, which can explain why many regimes that invest heavily in repression also air extreme propaganda claims (Gehlbach 2018).

Importantly, I assume that the citizen's stakes, Δ, are an increasing and (weakly) concave function of material incentives is, $\frac{\partial \Delta}{\partial r} > 0$ and $\frac{\partial^2 \Delta}{\partial r^2} \leq 0$. This can be justified by either assuming that the citizen obtains a reward when achieving the more favourable outcome $y = 1$ (which would imply that $u(1)$ is an increasing function of r), or by assuming that the citizen is punished when failing to achieve the outcome $y = 1$ (which would mean that $u(0)$ is a decreasing function of r). Finally, for consistency with the baseline model, I assume that there is scope of more demanding identity norms even if the level of material incentive is 0; that is, $\hat{e}_E < \frac{\Delta(0)}{c}$.[22]

Replicating the steps from above, we have that optimal citizen effort is given by $e^*(r, \hat{e}) = \frac{\Delta(r)+\alpha\hat{e}}{\alpha+c}$. This term is increasing both in the level of the identity norm $\hat{e}$ and in the size of the material incentives r. Moreover, the citizen affirms new identity content if:

$$\hat{e}_P \leq \frac{2\Delta(r)}{c} - \hat{e}_E.$$

Intuitively, an increase in material incentives increases the likelihood that the proposed identity is accepted and relaxes the affirmation constraint of the citizen.

Given the citizen's calculus, the propagandist's optimization problem is:

$$\max_{r,\hat{e}_P} \; 1\left(\hat{e}_P \leq \frac{2\Delta}{c} - \hat{e}_E\right)\left[\frac{\Delta(r) + \alpha\hat{e}_P}{\alpha + c}\right] + 1\left(\hat{e}_P > \frac{2\Delta}{c} - \hat{e}_E\right)\left[\frac{\Delta(r) + \alpha\hat{e}_E}{\alpha + c}\right] - \Psi(r)$$

[21] Specifically: $\frac{\partial \Delta}{\partial r}|_{r=0} > 0$, $\frac{\partial \Delta}{\partial r}|_{r=\bar{r}} < \infty$, $\lim_{r \to 0} \Psi'(r) = 0$, and $\lim_{r \to \bar{r}} \Psi'(r) = \infty$.

[22] This assumption requires that $\Delta(0) \geq 0$. Besides consistency with the baseline model, the assumption is made for analytical convenience. More generally, one could imagine that this inequality only holds for levels of r above a certain threshold, incentivizing the leader to choose levels of r above it.

her utility is increasing in $\hat{e}_P$ whenever $\hat{e}_P \leq \frac{2\Delta(r)}{c} - \hat{e}_E$ and independent of $\hat{e}_P$ otherwise. Consequently, it must be the case that $\hat{e}_P = \frac{2\Delta(r^*)}{c} - \hat{e}_E$. Plugging this expression into the maximization problem yields that the optimal level of material incentives, r^*, must be the positive (by Inada conditions) and unique (by concavity) solution to the following first-order condition:

$$(\alpha + c)^{-1} \frac{\partial \Delta}{\partial r} \left[1 + \frac{2\alpha}{c} \right] - \frac{\partial \Psi}{\partial r} = 0.$$

Examining the propagandist's optimal choices reveals the following:

Proposition 4. *When material incentives become (marginally) cheaper to impose, identity propaganda becomes more demanding.*

The analysis has two important implications. First, the spillover effect documented in Proposition 4 is consistent with the empirical regularity that regimes that most heavily invest in propaganda are often also the ones that repress their citizenry most severely (Gehlbach 2018). Examples include Nazi Germany, the Soviet Union, and North Korea today. Interestingly, however, current game-theoretic work frequently finds the opposite; regimes, or dictators, that rely heavily on repression (propaganda) do not use propaganda (repression). A general intuition for this relationship of substitutability can be described as follows: a dictator wishes to achieve some objective (for example, staying in office, successful mobilization) and can use material incentives as well as propaganda to obtain this goal. In equilibrium, the dictator will choose a higher level of the instrument that has the better effectiveness-cost ratio. In turn, if the costs for either material incentives or propaganda increase, the dictator will shift resources to the other instrument (Luo and Rozenas 2018; Tyson and Smith 2018; Wintrobe 1990). Relatedly, Guriev and Treisman (2015) find that (factual) propaganda and repression are substitutes; that is, the regime uses either propaganda or repression, but never both. The intuition is different, however. In their model, there is uncertainty over the competence of the dictator, and the use of repression reveals the dictator to be incompetent. As a result, when using repression, belief manipulation through propagandistic communication is completely ineffective and will not be employed by the dictator. Here, by contrast, the citizen does not learn about the leader's competence but, rather, affirms identity norms. This creates an endogenous source of complementarity; higher material incentives increase the citizen's stakes, rendering him more likely to affirm demanding identity norms to increase his expected payoff when mobilizing for political action.[23]

Second, the analysis suggests a rationale for effective identity propaganda even for the case in which citizens are initially materially opposed to exerting any effort, that is, $u(0) > u(1)$. As discussed above, when citizens value the outcome $y = 0$ more than outcome $y = 1$, identity propaganda cannot be effective. However, it might be the case that the propagandists can employ material incentives to render the stakes positive – which also enables effective identity propaganda. Thus, giving the propagandist the ability to cheaply allocate material resources may also give the propagandist the ability to effectively use propagandistic communication to shape behavior.

[23]Other scholarship also identifies sources of complementarity between material incentives and propaganda. However, these contributions do not focus on identity statements, and the complementarity is due to very different reasons. In Chen and Xu (2017), economic reforms and propaganda go hand in hand because material improvements increase the regime's credibility, which it exploits by airing more intense propaganda. In Horz (2021), repression and propagandistic communication, modelled as 'spins of events', are complements. There, the cause is that repression reduces the incentives of the audience to be Bayesian information processors, rendering it more willing to accept propaganda unquestioningly.

Conclusion

In this paper, I analyzed an important subset of propaganda which I call identity propaganda. When broadcasting identity propaganda, the propagandist attempts to change opinions about what it means to be a member of a social group. My analysis implies that propagandists are able to exploit their agenda-setting power to design effective identity norms, which explains a large set of empirical findings (Bleck and Michelitch 2017; DellaVigna et al. 2014; Enikolopov, Petrova, and Zhuravskaya 2011; Mitts, Phillips, and Walter 2022; Peisakhin and Rozenas 2018; Yanagizawa-Drott 2014). Moreover, when the propagandist's audience is heterogeneous in terms of their preferences, mobilization costs and pre-existing norms can have an ambiguous effect on identity norms because an increase in either factor causes group members to affirm less demanding norms, but propagandists respond by strategically designing divisive norms that are accepted by the more radical members of the group only. The fact populist leaders seem to use especially divisive rhetoric when attempting to mobilize previously marginalized groups (who presumably face high costs of participating in collective action) provides some support for this prediction. I also show that when the propagandists can cheaply allocate material rewards or punishments, identity propaganda will similarly be more demanding. By contrast, the nature of strategic interaction at the mobilization stage has an ambiguous effect on identity norms.

My analysis shows that psychological and rational choice approaches to politics can be fruitfully combined. Historically, these approaches have provided competing explanations for political phenomena. Recent work shows that they can be fruitfully combined to provide more convincing explanations for a range of phenomena (Acharya, Blackwell, and Sen 2018; Diermeier and Li 2019). In this paper, identity matters for effort because of non-material, psychological concerns. However, the extent to which identities can change depends on material concerns – as implied by the citizen's rational calculus. Thus, material and non-material considerations jointly shape political decisions and outcomes.

This paper also suggests several avenues for future work. First, I assume that the propagandist has the exclusive agenda-setting power to propose identity norms. By contrast, future work should explicitly endogenize the conditions under which social group leaders have the authority to put forward new identity norms. An empirically plausible scenario involves competition by another elite actor because, in all but the most dictatorial regimes, propagandists usually have to take into account communication by other elites. In the Appendix, I provide the first step in this direction by incorporating a second propagandist into the baseline model. I assume that the second propagandist has different material interests than the propagandist in the baseline case, and I show that equilibrium identity norms converge to the optimal identity norm described above. The reason is that both propagandists are incentivized to 'undercut' their respective claims to induce the citizen to accept their proposed identity norm. Consequently, in a competitive environment, identity propaganda is still effective at changing behavior but now does so in a way that benefits the citizen. Future work may extend this analysis to consider the effects of strategic interaction at the mobilization stage when propagandists compete for attention or to consider the effects of allowing one propagandist to partially censor the other.

Second, scholars could extend the model to consider the interaction between factual and identity propaganda, as well as the interaction between salience and norm identity propaganda. However, given the scarcity of formal work involving identity propaganda, I abstract away from these kinds of interactions in this paper and consider a special case in which the propagandist solely engages in identity-based norm propaganda. In practice, of course, propagandistic statements often contain aspects of both factual and non-factual (for example, identity), propaganda, or messaging about different identity norms and about increasing the salience of identity – even a single sentence can contain several types. Moreover, factual propaganda may be uttered with the aim of not only shaping the audience's perception of the state of the world but also shaping its

identity; propagandistic claims about whether or not massacres or other atrocities 'really happened' is one example. Scrutinizing such communication while taking into account receivers' incentives to affirm communication and hence change their attitudes is an important avenue for future work.[24]

Finally, future work could also conceptualize identity differently. There are two ways to augment my analysis. On the one hand, I focused on identity norms and relied on a particular functional form for the identity portion of the citizen's utility function, which featured symmetric costs for 'too low' and 'too high' effort choices. In contrast to this case, leaders sometimes communicate to citizens that they ought to at least make some level of effort, but if they want to do more, that is even better. In the Appendix, I provide a first step for a more complete future analysis by considering alternative specifications of the citizen's utility function. I briefly discuss the case in which $U^I = \alpha 1(e \geq \hat{e})$, which means that the citizen receives an identity 'bonus' (α) whenever his effort choice is at least equal to the active norm, $\hat{e}$. In this case, the conditions for effective identity propaganda are different – the propagandist is influential if existing norms are relatively demanding – but the comparative static results for the propagandist's choice are identical, increasing in the stakes Δ but decreasing in the mobilization costs c. Thus, while some of the results here rely on the symmetry of the loss function, some comparative static results are robust.

On the other hand, future work could also look beyond identity norms, scrutinizing the propagandist's ability to manipulate when identity is conceptualized differently. In the Appendix, I show how the general approach to identity propaganda – the leader makes a proposal about a new meaning of an identity while the citizen can affirm or reject it – can be adapted to the case in which identity matters because it implies a particular configuration of citizens' social preferences. For example, a citizen's identity can be more or less altruistic with respect to another citizen. These different action spaces and identity parameters yield richer interpretations – for example, identities can be 'antagonistic,' as in the ethnic politics literature – and potentially tighter connections to empirical analyses.[25]

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0007123423000182.

Acknowledgements. I thank Eric Dickson, Adam Przeworski, Alastair Smith, Catherine Hafer, Arturas Rozenas, Hannah Simpson, Andrew Little, Jeremy Wallace, Jessica Sun, Nadiya Kostyuk, and participants at MPSA 2016, APSA 2016, the Behavioral Models of Politics Conference 2016, and the Formal Models of International Relations Conference 2020 for helpful comments on earlier drafts of this paper.

Financial support. None.

Competing interests. None.

References

Acharya A, Blackwell M and Sen M (2018) Explaining preferences from behavior: A cognitive dissonance approach. *The Journal of Politics* **80**(2), 400–11.

Adena M et al. (2015) Radio and the rise of the Nazis in prewar Germany. *Quarterly Journal of Economics* **130**(4), 1885–1939.

Ahlquist JS and Levi M (2011) Leadership: What it means, what it does, and what we want to know about it. *Annual Review of Political Science* **14**, 1–24.

Akerlof GA and Kranton RE (2000) Economics and identity. *Quarterly Journal of Economics* **115**(3), 715–52.

[24]In the Appendix, I provide an initial analysis of identity propaganda that aims at changing an identity's salience, which is empirically common; see, for example, Brady (2012) for the case of China. I show that when citizens take statements about salience as given, the propagandist is incentivized to increase an identity's salience if this pushes another (material or group) identity to the background.

[25]One limitation of this approach is that for the specification considered in the Appendix, the propagandist's optimal choice is always a corner solution, which limits the usefulness of the model to explain variations in identity propaganda.

Akerlof GA and Kranton RE (2005) Identity and the economics of organizations. *Journal of Economic perspectives* **19**(1), 9–32.

Alonso R and Câmara O (2016) Bayesian Persuasion with heterogeneous priors. *Journal of Economic Theory* **165**, 672–706.

Amenta E et al. (2010) The political consequences of social movements. *Annual Review of Sociology* **36**, 287–307.

Anduiza E, Guinjoan M and Rico G (2019) Populism, participation, and political equality. *European Political Science Review* **11**(1), 109–24.

Askew M and Helbardt S (2012) Becoming Patani warriors: Individuals and the insurgent collective in Southern Thailand. *Studies in Conflict & Terrorism* **35**(11), 779–809.

Baker WD and Oneal JR (2001) Patriotism or opinion leadership? The nature and origins of the "Rally 'round the flag" effect. *Journal of Conflict Resolution* **45**(5), 661–87.

Baliga S and Sjöström T (2012) The strategy of manipulating conflict. *American Economic Review* **102**(6), 2897–2922.

Barber B and Miller C (2019) Propaganda and combat motivation: Radio broadcasts and German soldiers' performance in World War II. *World Politics* **71**(3), 457–502.

Bénabou R and Tirole J (2011) Identity, morals and taboos: Beliefs as assets. *The Quarterly Journal of Economics* **126**(2), 805–55.

Bleck J and Michelitch K (2017) Capturing the airwaves, capturing the nation? A field experiment on state-run media effects in the wake of a coup. *The Journal of Politics* **79**(3), 873–89.

Blouin A and Mukand SW (2019) Erasing ethnicity? Propaganda, nation building, and identity in Rwanda. *Journal of Political Economy* **127**(3), 1008–62.

Bokemper SE et al. (2022) Testing persuasive messaging to encourage COVID-19 risk reduction. *PLoS ONE* **17**(3), 1–20.

Brady A-M (2012) "We are all part of the same family": China's ethnic propaganda. *Journal of Current Chinese Affairs* **41**(4), 159–81.

Brass PR (1997) *Theft of an Idol: Text and Context in the Representation of Collective Violence*. Princeton, NJ: Princeton University Press.

Bueno De Mesquita E (2010) Regime change and revolutionary entrepreneurs. *American Political Science Review* **104**(03), 446–66.

Callander S and Wilkie S (2007) Lies, damned lies, and political campaigns. *Games and Economic Behavior* **60**, 262–86.

Carrillo JD and Gromb D (1999) On the strength of corporate cultures. *European Economic Review* **43**(4-6), 1021–37.

Carter EB and Carter BL (2021) Propaganda and protest in autocracies. *Journal of Conflict Resolution* **65**(5), 919–49.

Cheeseman N and Larmer M (2015) Ethnopopulism in Africa: Opposition mobilization in diverse and unequal societies. *Democratization* **22**(1), 22–50.

Chen Y and Li SX (2009) Group identity and social preferences. *American Economic Review* **99**(1), 431–57.

Chen J and Xu Y (2017) Information manipulation and reforms in authoritarian regimes. *Political Science Research and Methods* **5**(1), 163–78.

de la Torre C (2017) Populism in Latin America. In Kaltwasser CR, Taggart P, Espejo PO and Ostiguy P (eds), *The Oxford Handbook of Populism*. Oxford, England: Oxford University Press, pp. 195–213.

DellaVigna S and Gentzkow M (2010) Persuasion: empirical evidence. *Annual Review of Economics* **2**(1), 643–69.

DellaVigna S et al. (2014) Cross-border media and nationalism: Evidence from Serbian radio in Croatia. *American Economic Journal: Applied Economics* **6**(3), 103–32.

Dickson ES (2010) Leadership, Followership, and Beliefs About the World: An Experiment. *Working Paper*.

Dickson ES and Scheve K (2006) Social identity, political speech, and electoral competition. *Journal of Theoretical Politics* **18**(1), 5–39.

Dickson ES and Scheve K (2010) Social identity, electoral institutions and the number of candidates. *British Journal of Political Science* **40**(2), 349–75.

Diermeier D and Li C (2019) Partisan affect and elite polarization. *American Political Science Review* **113**(1), 277–81.

Edmond C (2013) Information manipulation, coordination, and regime change. *The Review of Economic Studies* **80**(4), 1422–58.

Eifert B, Miguel E and Posner DN (2010) Political competition and ethnic identification in Africa. *American Journal of Political Science* **54**(2), 494–510.

Enikolopov R, Petrova M and Zhuravskaya E (2011) Media and political persuasion: evidence from Russia. *American Economic Review* **101**(7), 3253–85.

Fearon JD and Laitin DD (2000) Violence and the social construction of ethnic identity. *International Organization* **54**(4), 845–77.

Gehlbach S (2018) What is next for the study of non-democracy?. In Ménard C and Shirley MM (eds), *A Research Agenda for New Institutional Economics*. Northampton, MA, USA: Edward Elgar Publishing.

Gehlbach S and Sonin K (2014) Government control of the Media. *Journal of Public Economics* **118**, 163–71.

Gehlbach S, Sonin K and Svolik MW (2016) Formal models of nondemocratic politics. *Annual Review of Political Science* **19**, 565–84.

Gibbons R (2010) Inside organizations: Pricing, politics, and path dependence. *Annual Review of Economics* **2**(1), 337–65.

Gibilisco M, Kenkel B and Rueda MR (2022) Competition and civilian victimization. *Journal of Conflict Resolution* **66**(4-5), 809–35.
Glaeser EL (2005) The political economy of hatred. *Quarterly Journal of Economics* **120**(1), 45–86.
Grossman G et al. (2020) Political partisanship influences behavioral responses to governors' recommendations for COVID-19 prevention in the United States. *Proceedings of the National Academy of Sciences* **117**(39), 24144–153.
Guriev S and Treisman D (2015) How modern dictators survive: Cooptation, censorship, propaganda, and repression. *NBER Working Paper* (w21136), 1–36.
Hallsworth M et al. (2017) The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics* **148**, 14–31.
Horowitz D (1985) *Ethnic Groups in Conflict*. Berkeley, CA, USA: University of California Press.
Horz CM (2021) Propaganda and skepticism. *American Journal of Political Science* **65**(3), 717–732.
Huang H (2015) Propaganda as signaling. *Comparative Politics* **47**(4), 419–37.
Huddy L (2001) From social to political identity: A critical examination of social identity theory. *Political Psychology* **22**(1), 127–56.
Humphreys M and Weinstein JM (2006) Handling and manhandling civilians in civil war. *American Political Science Review* **100**(3), 429–47.
Invernizzi GM et al. (2021) Tra i Leoni: Revealing the preferences behind a superstition. *Journal of Economic Psychology* **82**, 1–11.
Jowett GS and O'Donnell VJ (2015) *Propaganda and Persuasion*, 6th edn. Los Angeles, CA, USA: Sage Publications.
Jun SJ and Lee S (2018) Identifying the effect of persuasion. *arXiv preprint arXiv:1812.02276*.
Kalin M and Sambanis N (2018) How to think about social identity. *Annual Review of Political Science* **21**, 239–57.
Kapferer B (2011) *Legends of People, Myths of State: Violence, Intolerance, and Political Culture in Sri Lanka and Australia*. New York City, NY, USA: Berghahn Books.
Kershaw I (2001) *The "Hitler Myth:" Image and Reality in the Third Reich*. Oxford, England: Oxford University Press.
Kobayashi T and Katagiri A (2018) The "Rally 'round the flag" effect in territorial disputes: Experimental evidence from Japan–China relations. *Journal of East Asian Studies* **18**(3), 299–319.
Landa D and Duell D (2015) Social identity and electoral accountability. *American Journal of Political Science* **59**(3), 671–89.
Landa D and Tyson SA (2017) Coercive leadership. *American Journal of Political Science* **61**(3), 559–74.
Lehmann TC and Tyson SA (2022) Sowing the seeds: Radicalization as a political tool. *American Journal of Political Science* **66**(2), 485–500.
Little AT (2017) Propaganda and credulity. *Games and Economic Behavior* **102** (March), 224–32.
Luo Z and Rozenas A (2018) Strategies of election rigging: Trade-offs, determinants, and consequences. *Quarterly Journal of Political Science* **13**(1), 1–28.
McCoy J and Somer M (2019) Toward a theory of pernicious polarization and how it harms democracies: Comparative evidence and possible remedies. *The ANNALS of the American Academy of Political and Social Science* **681**(1), 234–71.
Mitts T, Phillips G and Walter BF (2022) Studying the impact of ISIS propaganda campaigns. *The Journal of Politics* **84**(2), 1220–25.
Mukand S and Rodrik D (2018) The political economy of ideas: On ideas versus interests in policymaking. *National Bureau of Economic Research* Working Paper **24467**, 1–42. Available at https://www.nber.org/papers/w24467.
Newman B et al. (2021) The Trump effect: An experimental investigation of the emboldening effect of racially inflammatory elite communication. *British Journal of Political Science* **51**(3), 1138–59.
NPR (2021) Read Trump's Jan. 6 Speech, A Key Part of Impeachment Trial. Available at https://www.npr.org/2021/02/10/966396848/read-trumps-jan-6-speech-a-key-part-of-impeachment-trial.
Peisakhin L and Rozenas A (2018) Electoral effects of biased media: Russian television in Ukraine. *American journal of political science* **62**(3), 535–50.
Penn EM (2008) Citizenship versus ethnicity: The role of institutions in shaping identity choice. *Journal of Politics* **70**(4), 956–73.
Pepinsky T (2020) Migrants, minorities, and populism in Southeast Asia. *Pacific Affairs* **93**(3), 593–610.
Raymond L, Kelly D and Hennes EP (2021) Norm-based governance for severe collective action problems: lessons from climate change and COVID-19. *Perspectives on Politics*, 1–14.
Reno RR, Cialdini RB and Kallgren CA (1993) The transsituational influence of social norms. *Journal of Personality and Social Psychology* **64**(1), 104.
Sambanis N and Shayo M (2013) Social identification and ethnic conflict. *American Political Science Review* **107**(2), 294–325.
Schnakenberg KE (2014) Group identity and symbolic political behavior. *Quarterly Journal of Political Science* **9**(2), 137–67.
Shadmehr M and Bernhardt D (2019) Vanguards in revolution. *Games and Economic Behavior* **115**, 146–66.
Shayo M (2009) A model of social identity with an application to political economy: Nation, class, and redistribution. *American Political Science Review* **103**(02), 147.
Shayo M (2020) Social identity and economic policy. *Annual Review of Economics* **12**, 355–89.
Torun D and Myatt DP (2007) Leading the party: Coordination, direction, and communication. *American Political Science Review* **101**(4), 827–45.

Tyson SA and Smith A (2018) Dual-layered coordination and political instability: Repression, co-optation, and the role of information. *The Journal of Politics* **80**(1), 44–58.

Welch D (2014) *Nazi Propaganda (RLE Nazi Germany & Holocaust): The Power and the Limitations*. Milton Park, England: Routledge.

Wintrobe R (1990) The tinpot and the totalitarian: An economic theory of dictatorship. *American political science review* **84**(3), 849–72.

Woodward SL (1995) *Balkan Tragedy: Chaos and Dissolution After the Cold War*. Washington, DC, USA: Brookings Institution Press.

Yanagizawa-Drott D (2014) Propaganda and conflict: Theory and evidence from the Rwandan genocide. *Quarterly Journal of Economics* **129**(4), 1947–94.

Cite this article: Horz CM (2024). Identity Propaganda. *British Journal of Political Science* **54**, 313–338. https://doi.org/10.1017/S0007123423000182