Online Appendix: Identity Propaganda

Carlo M. Horz*

Contents

*Assistant Professor, Texas A&M University, Department of Political Science, College Station, TX 77845-4348 E-mail: carlo.horz@tamu.edu. Web: www.carlohorz.com.

1 Baseline Model

1.1 Omitted Proofs

Proof of Lemma 1

Proof. Plugging $e^*(\hat{e})$ into $\alpha\left(e^*(\hat{e}) - \hat{e}\right)$ and asking when the resulting quantity is greater than 0 yields:

$$\alpha\left(\frac{\Delta + \alpha\hat{e}}{\alpha + c} - \hat{e}\right) > 0$$

Re-arranging:

$$\Delta - \hat{e}c > 0$$

which is positive if $\hat{e} < \frac{\Delta}{c}$, negative if $\hat{e} > \frac{\Delta}{c}$, and equal to 0 if $\hat{e} = \frac{\Delta}{c}$. □

Proof of Lemma 2:

Proof. Define $e_1^* \equiv e^*(\hat{e}_P)$ and $e_0^* \equiv e^*(\hat{e}_E)$ as the effort levels when $t = 1$ and when $t = 0$, respectively. Then:

$$V(\hat{e}_P) - V(\hat{e}_E) = e_1^*\Delta - \frac{c}{2}\left(e_1^*\right)^2 - \frac{\alpha}{2}\left(e_1^* - \hat{e}_P\right)^2 - e_0^*\Delta - \frac{c}{2}\left(e_0^*\right)^2 - \frac{\alpha}{2}\left(e_0^* - \hat{e}_E\right)^2$$

or

$$\Delta(e_1^* - e_0^*) - \frac{c}{2}\left(e_1^* - e_0^*\right)\left(e_1^* + e_0^*\right) - \frac{\alpha}{2}\left[\left(e_1^* - \hat{e}_P\right)^2 - \left(e_0^* - \hat{e}_E\right)^2\right] \tag{1}$$

Defining $E_0 \equiv \frac{\Delta}{\alpha+c}$ and $E_1 \equiv \frac{\alpha}{\alpha+c}$ allows us to write $e^* = E_0 + E_1\hat{e}$. Using this notation, we have that expression 1 is equal to $(\hat{e}_P - \hat{e}_E)$ times

$$\Delta E_1 - \frac{c}{2}E_1\left(2E_0 + E_1\left(\hat{e}_P + \hat{e}_E\right)\right) - \frac{\alpha}{2}\left[2E_0(E_1 - 1) + (E_1 - 1)^2(\hat{e}_P + \hat{e}_E)\right]$$

Suppose for the time being that $\hat{e}_P \geq \hat{e}_E$. Consequently, we have that $V(\hat{e}_P) - V(\hat{e}_E) \geq 0$ if:

$$\Delta E_1 - cE_0E_1 - \alpha E_0(E_1 - 1) \geq (\hat{e}_P + \hat{e}_E)\left[\frac{c}{2}E_1^2 + \frac{\alpha}{2}\left(E_1 - 1\right)^2\right]$$

Plugging in E_0 and E_1 again simplifies the left-hand side to: $\frac{\alpha\Delta}{\alpha+c}$. The term inside the square brackets on the right-hand side simplifies to $\frac{\alpha c}{2(\alpha+c)}$. Consequently:

$$\frac{\alpha\Delta}{\alpha+c} \geq (\hat{e}_P + \hat{e}_E)\frac{\alpha c}{2(\alpha+c)}$$

Thus:

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

Alternatively, note that $V(\hat{e}_P) - V(\hat{e}_E)$ is a quadratic function of $\hat{e}_P$ (because U is quadratic in effort, e^* is linear in identity norms, and the only interaction is via $(e^*(\hat{e}) - \hat{e})^2$). Consequently, there must be two roots, one of which must be $\hat{e}_E$. By the symmetry of a polynomial of degree two around the maximizer of the function—which by Lemma 1 is $\frac{\Delta}{c}$, the second root must be equally far away from $\frac{\Delta}{c}$ as $\hat{e}_E$ is from $\frac{\Delta}{c}$. This implies that the second root is $\frac{2\Delta}{c} - \hat{e}_E$. □

Proof of Proposition 1:

Proof. Follows by the argument in the text and by simple inspection from $\hat{e}_P^* = \frac{2\Delta}{c} - \hat{e}_E$. □

1.2 Can Citizens Construct Identity Norms?

In the main text, I assume that the propagandist has exclusive agenda-setting rights. As a consequence, the model predicts that whenever citizens are "held back" by the existing norm ($\hat{e}_E < \frac{\Delta}{c}$), the propagandist can powerfully affect behavior by pushing for identity norms that are too demanding (relative to the optimal identity norm). Why does the propagandist have exclusive agenda-setting power? One justification is that constructing identity norms is a very difficult, costly process since a group's history, myths, and experiences have to be linked a specific behavioral prescription. Thus, not all citizens may be able or willing to engage in identity construction. To see this formally, suppose that the citizen can choose to construct his own identity norm or not. Given the analysis conducted above, it is apparent that if the citizen chooses to construct his own identity norm, the citizen chooses $\hat{e}_{\text{opt}}$—which completely *inoculates* the citizen against any identity propaganda attempts. By contrast, when not choosing to engage in identity construction himself, the outcome is $\hat{e}^*_P$. The citizen thus chooses not to choose an identity norm for himself if:

$$V(\hat{e}^*_P) \geq V(\hat{e}_{\text{opt}}) - \kappa \Rightarrow \kappa \geq V(\hat{e}_{\text{opt}}) - V(\hat{e}^*_P)$$

Thus, the analysis remains valid for citizens that face relatively high costs of designing identity norms for themselves.

1.3 Membership in Multiple Groups

Almost all citizens are members of multiple social groups. For example, a citizen can be a member of a nation, a class, and a religious group. Moreover, there is evidence that membership in one group can affect citizens' susceptibility of identity propaganda. For example, Spenkuch and Tillmann (2018) find that membership in the Catholic church decreases the likelihood of voting for the Nazi party in Weimar Germany. In this section, I examine the consequences of multiple identities and show that membership in other groups does indeed affect the propagandist's ability to push for more demanding identity norms by tightening or loosening the citizen's affirmation constraint.

Formally, suppose there are two relevant social groups, A and B. Consequently, the citizen has two identities, A and B, with identity norms $\hat{e}_A$ and $\hat{e}_B$ as well as saliency parameters α and β, respectively. The propagandist can only propose a new identity norm for group A, i.e., $\hat{e}_A$, while the norm associated with membership in B, $\hat{e}_B$, remains the same. The citizen's total utility function is:

$$U = \underbrace{eu(1) + (1-e)u(0) - \frac{c}{2}e^2}_{=U^M} - \frac{\alpha}{2}\underbrace{(e - \hat{e}_A)^2}_{=U^I_A} - \frac{\beta}{2}\underbrace{(e - \hat{e}_B)^2}_{=U^I_B}$$

Maximizing this utility function with respect to effort e yields the optimal (interior) effort choice given identity norms for groups A and B:

$$e^*(\hat{e}_A) = \frac{\Delta + \alpha \hat{e}_A + \beta \hat{e}_B}{\alpha + \beta + c}$$

Plugging this expression back into the citizen's utility function yields the equilibrium utility as a function of the identity norm $\hat{e}_A$: $V(\hat{e}_A) = U(e^*(\hat{e}_A), \hat{e}_A)$. The next result examines the behavior of this function:

Lemma 1. *The function $V(\hat{e}_A)$ has the following properties:*

- *$V(\hat{e}_A)$ is increasing in $\hat{e}_A$ if $\hat{e}_A < \frac{\Delta+\beta\hat{e}_B}{\beta+c}$, decreasing in $\hat{e}_A$ if $\hat{e}_A > \frac{\Delta+\beta\hat{e}_B}{\beta+c}$, and equal to 0 if $\hat{e}_A = \frac{\Delta+\beta\hat{e}_B}{\beta+c}$.*
- *The citizen's optimal identity norm is $\hat{e}_{A,opt} = \frac{\Delta+\beta\hat{e}_B}{\beta+c}$*
- *If $\hat{e}_{AE} < \hat{e}_{A,opt}$, the citizen affirms new identity content if:*

$$\hat{e}_{AP} \in \left[\hat{e}_{AE}, \frac{2(\Delta + \beta\hat{e}_B)}{\beta + c} - \hat{e}_{AE}\right]$$

Proof. I proceed to prove the three parts separately:

Part 1 Differentiating $V(\hat{e}_A)$ with respect to $\hat{e}_A$ yields:

$$\alpha(e^*(\hat{e}_A) - \hat{e}_A) = \alpha\left(\frac{\Delta + \alpha\hat{e}_A + \beta\hat{e}_B}{\alpha + \beta + c} - \hat{e}_A\right)$$

This is positive if $\hat{e}_A < \frac{\Delta+\beta\hat{e}_B}{\beta+c}$, negative if $\hat{e}_A > \frac{\Delta+\beta\hat{e}_B}{\beta+c}$, and equal to 0 if $\hat{e}_A = \frac{\Delta+\beta\hat{e}_B}{\beta+c}$.

Part 2: By the calculations in part 1, $\frac{\partial V}{\partial \hat{e}_A} = 0 \Rightarrow e_{A,\text{opt}} = \frac{\Delta+\beta\hat{e}_B}{\beta+c}$.

Part 3: Write the optimal effort choice as:

$$e^*(\hat{e}_A) = \underbrace{\frac{\Delta + \beta\hat{e}_B}{c + \alpha + \beta}}_{\equiv E_0} + \underbrace{\frac{\alpha}{c + \alpha + \beta}}_{\equiv E_1}\hat{e}_A$$

Moreover, define $e_1^* \equiv e^*(\hat{e}_{AP})$ and $e_0^* \equiv e^*(\hat{e}_{AE})$, fot $t = 0, 1$. Then $V(\hat{e}_{AP}) \geq V(\hat{e}_{AE})$

can be written as follows:

$$e_1^*\Delta-\frac{c}{2}(e_1^*)^2-\frac{\alpha}{2}\left(e_1^*-\hat{e}_{AP}\right)^2-\frac{\beta}{2}\left(e_1^*-\hat{e}_B\right)^2 \geq e_0^*\Delta-\frac{c}{2}(e_0^*)^2-\frac{\alpha}{2}\left(e_0^*-\hat{e}_{AE}\right)^2-\frac{\beta}{2}\left(e_0^*-\hat{e}_B\right)^2$$

or

$$\Delta(e_1^*-e_0^*)-\frac{c}{2}(e_1^*-e_0^*)(e_1^*+e_0^*)-\frac{\alpha}{2}\underbrace{\left[(e_1^*-\hat{e}_{AP})^2-(e_0^*-\hat{e}_{AE})^2\right]}_{(1)}-\frac{\beta}{2}\underbrace{\left[(e_1^*-\hat{e}_B)^2-(e_0^*-\hat{e}_B)^2\right]}_{(2)}\geq 0 \tag{2}$$

Consider expression (1) in 2:

$$\begin{aligned}&2E_0(E_1-1)\hat{e}_{AP}+(E_1-1)^2\hat{e}_{AP}^2-2E_0(E_1-1)\hat{e}_{AE}+(E_1-1)^2\hat{e}_{AE}^2\\=&(\hat{e}_{AP}-\hat{e}_{AE})\left[2E_0(E_1-1)+(E_1-1)^2(\hat{e}_{AP}+\hat{e}_{AE})\right]\end{aligned}$$

Now consider expression (2) in 2:

$$\begin{aligned}&(e_1^*)^2-2\hat{e}_Be_1^*+\hat{e}_B^2-\left((e_0^*)^2-2\hat{e}_Be_0^*+\hat{e}_B^2\right)\\=&(e_1^*-e_0^*)(e_1^*-e_0^*-2\hat{e}_B)\\=&E_1(\hat{e}_{AP}-\hat{e}_{AE})(2E_0+E_1(\hat{e}_{AP}+\hat{e}_{AE}-2\hat{e}_B))\end{aligned}$$

Consequently, expression 2 can be written as:

$$\begin{aligned}\Delta E_1(\hat{e}_{AP}-\hat{e}_{AE})-&\frac{c}{2}\left[E_1(\hat{e}_{AP}-\hat{e}_{AE})(2E_0+E_1(\hat{e}_{AP}+\hat{e}_{AE}))\right]\\&-\frac{\alpha}{2}(\hat{e}_{AP}-\hat{e}_{AE})\left[2E_0(E_1-1)+(E_1-1)^2(\hat{e}_{AP}+\hat{e}_{AE})\right]\\&-\frac{\beta}{2}E_1(\hat{e}_{AP}-\hat{e}_{AE})(2E_0+E_1(\hat{e}_{AP}+\hat{e}_{AE}-2\hat{e}_B))\geq 0\end{aligned}$$

Assuming that $\hat{e}_{AP}-\hat{e}_{AE}\geq 0$, one can divide by $\hat{e}_{AP}-\hat{e}_{AE}$ to obtain:

$$\Delta E_1-cE_0E_1-\beta E_0E_1+\beta E_1\hat{e}_B-\alpha E_0(E_1-1)\geq(\hat{e}_{AP}+\hat{e}_{AE})\left[\frac{c}{2}E_1^2+\frac{\beta}{2}E_1^2+\frac{\alpha}{2}\left(E_1-1\right)^2\right]$$

Plugging in E_1 and E_0 again, the left-hand side of the preceding inequality simplifies to

$\frac{\alpha(\Delta+\beta\hat{e}_B}{\alpha+\beta+c}$. Moreover, the right-hand side of the preceding inequality simplifies to $(\hat{e}_{AP} + \hat{e}_{AE})\frac{\alpha(c+\beta)}{2(\alpha+\beta+c)}$. Consequently, we have:

$$\frac{\alpha(\Delta + \beta\hat{e}_B}{\alpha + \beta + c} \geq (\hat{e}_{AP} + \hat{e}_{AE})\frac{\alpha(c + \beta)}{2(\alpha + \beta + c)}$$

or:

$$2\frac{\Delta + \beta\hat{e}_B}{\alpha + \beta + c} \geq \hat{e}_{AP} + \hat{e}_{AE} \Rightarrow \hat{e}_{AP} \leq 2\frac{\Delta + \beta\hat{e}_B}{\alpha + \beta + c} - \hat{e}_{AE}$$

□

Lemma 1 generalizes previous results to the case of multiple politically relevant identities. Optimal identity content for one group is now affected by existing norms implied by membership in other groups. This means that not only material incentives constrain the propagandist, but other identities as well. Specifically, if members of group B ought to exert a low level of effort ($\hat{e}_B$ small), then the citizen's affirmation strategy implies that acceptable content for group A must be relatively low as well. Conversely, if members of group B ought to exert a high level of effort ($\hat{e}_B$ large), then acceptable content for group A can also be higher.

While identity norms implied by membership in group B affect the citizen's affirmation strategy, the qualitative dynamic of the propagandist-citizen interaction remains in place: since the propagandist still wishes to instill the most demanding identity norm that is affirmed, she proposed the following identity norm:

$$\hat{e}_A^* = \frac{2\,(\Delta + \beta\hat{e}_B)}{\beta + c} - \hat{e}_{AE}.$$

By inspection, this is increasing in the norm espoused by the other social group, $\hat{e}_B$: the more demanding identity norms in other groups, the more demanding identity propaganda will be for group A. This is consistent with existing work examining behavior (e.g., Spenkuch and Tillmann, 2018), but the more direct implication regarding propaganda remains yet to be tested.

2 Alternative Identity Utility Functions

In the main text, I focused on a particular functional form for identity-portion of the citizen's utility function: $U^I = -\frac{\alpha}{2}(e - \hat{e})^2$. A natural question is how the paper's conclusion change as different functional forms are considered. Here, I provide additional analyses to address this question.

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); Shayo (2009) to perceived distance to group ideal types (Shayo, 2009) and prescribed actions (Akerlof and Kranton, 2000). The general methodology of this paper is to describe identity propaganda is a bargaining process: the propagandist proposes a U_I while the citizen can either affirm (accept) or not affirm (reject) it. A simple way to accomplish this is to assume that the propagandist chooses a one-dimensional parameter, such as $\hat{e}_P \in [0, 1]$, and the citizen chooses to accept it, changing the utility function to $-(e - \hat{e}_P)^2$, or reject it, keeping the existing norm $\hat{e}_E$ and associated utility function $-(e - \hat{e}_E)^2$. In general, there should be a tension between the citizen's material interest, U^M and identity concerns, U^I. Using the specification of U^M stated in the main text, $e^* = \frac{\Delta}{c}$ maximizes the citizen's material interests. If $\hat{e} \neq e^*$, there is a tension between material and identity concerns. If identity concerns are endogenous, the best what can be achieved is a complete alignment of these two different concerns.

To probe this intuition further, consider first a general utility function $U^I(e, \hat{e})$. For a model that seeks to explain variation in identity propaganda's content and effectiveness, two conditions have to be satisfied. First, it needs to be the case that different identity conceptions induce different optimal (best response) effort levels. Formally, the effectiveness of effort varies with different substantive definitions of social identity. Assuming U^I is twice-differentiable, we require:

$$\frac{\partial e^*}{\partial \hat{e}} \neq 0 \quad \text{if and only if} \quad \frac{\partial^2 U^I}{\partial e \partial \hat{e}} \neq 0.$$

If this condition is not satisfied, optimal effort does not vary with the identity norm.

Given that the propagandist is only interested in shaping effort, any identity norm would then be optimal, from her perspective.

Relatedly, as discussed in the main text, the "scale" of U^I is not important. Because I assume that the affirmation choice is governed by the difference in equilibrium utilities, $V(\hat{e}_P) - V(\hat{e}_E)$, adding a constant to U^I does not affect the analysis. For example, in the main text, we have that $U^I = -\frac{\alpha}{2}(e - \hat{e})^2 \leq 0$, but I could alternatively employ

$$U_I = \frac{\alpha}{2}\left[1 - (e - \hat{e})^2\right].$$

Then, $U_I \geq 0$ for all e but the analysis goes through as before.

Second, we need that the derivative of the value function $V(\hat{e}) = U(e^*(\hat{e}), \hat{e})$ with respect to $\hat{e}$ is sometimes positive and sometimes negative, i.e., $\frac{\partial U^I}{\partial \hat{e}}\Big|_{e=e^*(\hat{e})} \lesseqgtr 0$. Substantively, this means that conditional on the actor behaving optimally, changes in identity conceptions can induce both positive and negative changes in the actor's psychological utility. These two conditions are intuitively plausible and can be readily checked in future applications.

These conditions rule out some choices for U^I that actually appear substantively not implausible. For example,

$$U^I(e, \hat{e}) = f(e) - \hat{e},$$

for an increasing function f, such as $f(e) = e$ or $f(e) = \sqrt{e}$, is ruled out by the first condition.

Below, I consider several other candidates for U^I. I discuss equilibrium intuitions and relate them to the baseline analysis.

2.1 Must do At Least This

Suppose that the identity-portion of the citizen's utility function is given by the following:

$$U^I = \alpha \mathbb{1}(e \geq \hat{e}).$$

Substantively, it means that the citizen receives α is their level of effort is above the threshold set by the identity norm, $\hat{e}$. One can think of the norm $\hat{e}$ as setting the threshold for acceptable group behavior.

The citizen's total utility function is:

$$\begin{aligned} U &= U^M + U^I \\ &= eu(1) + (1-e)u(0) - \frac{c}{2}e^2 + \alpha \mathbb{1}\left(e \geq \hat{e}\right) \end{aligned}$$

All other aspects of the game remain unchanged.

Turning to the equilibrium analysis, consider an arbitrary norm $\hat{e}$ and maximize U. To find the optimal level of effort, recall that $e = \frac{\Delta}{c}$ maximizes U^M. Thus, there are two candidates for an optimal choice, $\frac{\Delta}{c}$ or $\hat{e}$ (the minimal level of effort to secure the identity payoff α). There are several several cases to consider, depending on the ordering of these two candidates:

When the identity norm is relatively permissive, $\hat{e} < \frac{\Delta}{c}$, then the citizen optimally chooses $\frac{\Delta}{c}$. The equilibrium utility in this case is:

$$\frac{\Delta}{c}\Delta + u(0) - \frac{c}{2}\left(\frac{\Delta}{c}\right)^2 + \alpha.$$

When the identity norm is relatively demanding, $\hat{e} \geq \frac{\Delta}{c}$, then the citizen optimally chooses $\hat{e}$ if

$$\hat{e} \leq \frac{\Delta}{c} + \sqrt{\frac{2\alpha}{c}}.$$

The equilibrium utility in this case is:

$$\hat{e}\Delta + u(0) - \frac{c}{2}\left(\hat{e}\right)^2 + \alpha$$

Finally, if $\hat{e} > \frac{\Delta}{c} + \sqrt{\frac{2\alpha}{c}}$, then the citizen optimally chooses $\frac{\Delta}{c}$, and his equilibrium utility is:

$$\frac{\Delta}{c}\Delta - \frac{c}{2}\left(\frac{\Delta}{c}\right)^2 + u(0).$$

Summarizing these cases, we have:

- If the norm the citizen adheres to is very permissive, the citizen chooses the action that maximizes his material utility while also enjoying identity benefits.
- If the norm is intermediate, the citizens acts according to it which has a material loss but an identity benefit.
- If the norm is very demanding, the citizen foregoes it and simply maximizes his material utility.

Now consider endogenous norms. As before, denote by $\hat{e}_E$ the existing norm. There are three cases:

1. The existing norm is relatively less demanding, i.e., $\hat{e}_E < \frac{\Delta}{c}$. Here, the citizen never accepts new identity content. As a result, the propagandist can chooses any $\hat{e}_P \in [0, 1]$.
2. The existing norm is in an intermediate range, i.e., $\hat{e}_E \in \left(\frac{\Delta}{c}, \frac{\Delta}{c} + \sqrt{\frac{2\alpha}{c}}\right)$. The citizen affirms any $\hat{e}_P \leq \frac{2\Delta}{c} - \hat{e}_E$. But, since any accepted new identity norm is lower than the existing norm (the most demanding identity norm is $\frac{\Delta}{c}$ which is still smaller than the existing one), so the propagandist will never propose any norm that the citizen affirms. Hence, the propagandist chooses a norm that will not be affirmed.
3. The existing norm is relatively demanding, i.e., $\hat{e}_E \geq \frac{\Delta}{c} + \sqrt{\frac{2\alpha}{c}}$. Here, the citizen accepts if $\hat{e}_P \leq \frac{\Delta}{c} + \sqrt{\frac{2\alpha}{c}}$. The propagandist chooses the most demanding norm that is affirmed, so the equilibrium choice is $\hat{e}_P^* = \frac{\Delta}{c} + \sqrt{\frac{2\alpha}{c}}$.

Clearly, the analysis is less tractable that the baseline specification with $U^I = -\frac{\alpha}{2}(e - \hat{e})^2$. Substantively, compared to the baseline analysis in the main text, the key difference is that here, the propagandist is only influential if existing norms are too extreme (so they are ignored by the citizen). By contrast, in the main text, the propagandist is influential if identity norms are permissive ($\hat{e}_E < \frac{\Delta}{c}$).

However, it is important to emphasize that the comparative static results of the optimal norm are quite similar to the case analyzed in the main text. In particular, when influential, $\hat{e}_P^*$ is increasing in Δ and decreasing in mobilization costs c. This provides a degree of robustness to the analysis conducted in the main text.

2.2 Loss Up Until a Point

Now assume that the identity-portion of the citizen's utility function is given by:

$$U^I = \begin{cases} -\frac{\alpha}{2}\left(e-\hat{e}\right)^2 & \text{if} e \leq \hat{e} \\ 0 & \text{otherwise.} \end{cases}$$

The citizen's total utility function is thus:

$$U = U^M + U^I = eu(1) + (1-e)u(0) - \frac{c}{2}e^2 + U_I$$

To find the citizen's optimal level of effort, one has to consider two cases:

Suppose first that $e \leq \hat{e}$ in equilibrium. The citizens optimal choice is $e^* = \frac{\Delta+\alpha\hat{e}}{\alpha+c}$. For consistency, we require that $e^* \leq \hat{e}$, which requires $\frac{\Delta}{c} \leq \hat{e}$.

Now suppose that $e > \hat{e}$ in equilibrium. The citizen's optimal choice is $e^* = \frac{\Delta}{c}$, which for consistency requires $\frac{\Delta}{c} > \hat{e}$.

Thus, for permissive identity norms such that $\hat{e} \leq \frac{\Delta}{c}$, the citizen's optimal choice is $e^* = \frac{\Delta+\alpha\hat{e}}{\alpha+c}$. Conversely, for extreme identity norms that such $\hat{e} > \frac{\Delta}{c}$, the citizen's optimal choice is $\frac{\Delta}{c}$.

Now consider the value of an identity norm, $V(e^*(\hat{e}), \hat{e})$.

In the first case, the derivative of V with respect to $\hat{e}$ is always negative. (It is positive if $\hat{e} < \frac{\Delta}{c}$, but this case is not possible because identity norms must be relatively demanding here.) Thus, the citizen will never accept a more demanding identity norm. So the propagandist will simply propose a norm that will not be affirmed.

In the second case, the derivative of V with respect to $\hat{e}$ is 0. Thus, there is never an incentive to accept any identity change, as long as $\hat{e} < \frac{\Delta}{c}$.

The analysis spells out an important boundary condition for the results derived in the main text. When the citizen faces some costs for doing "too little," it has to be the case that the citizen faces face some cost for doing "too much."[1] This requirement is plausible for some cases but not so in others. For example, interpreting U^I broadly, "overachievers" sometimes face peer punishment and associated negative psychological emotions. By contrast, they are also sometimes seen as role models and may internalize associated positive emotions. Thus, the applicability of the model depends on the intended applications and to what extent each case is plausible.

2.3 Beyond Norms: Identity and Social Preferences

Here, I briefly discuss the extent to which my model can incorporate an alternative conceptualizations of identity based on social preferences (e.g., Shayo, 2009; Bénabou and Tirole, 2011; Chen and Li, 2009). To distinguish the analysis from my treatment of identity norms, I focus on a different one-dimensional parameter. Thus, rather than focus on $\hat{e} \in [0,1]$, I will focus on the value of $\tau \in [-1,1]$. Substantively, a negative value of τ will indicate that the citizen is "spiteful" whereas a positive value of τ will indicate "altruistic" attitudes. This is related to the notion of antagonistic vs. non-antagonistic identities in the literature on ethnic conflict (Fearon and Laitin, 2000).

Specifically, consider the variant of the model in which there are two citizens, a and b, and the probability of obtaining outcome $y = 1$ is given by $\gamma(e_a, e_b) = \frac{e_a + e_b}{2}$. However, different from above, assume b's effort choice is exogenously fixed at $\tilde{e}_b$, and b does not have any identity concerns.[2] Rather than facing behavioral prescriptions, citizen a now has *social preferences* so that the identity-based portion of his utility function is given by:

$$U_a^I = \tau U_b^M$$

where $\tau \in [-1,1]$. Thus, τ determines whether b's welfare enters positively or negatively into a's utility.

[1]Note that when $U^I = \alpha \mathbb{1}(e \geq \hat{e})$, the citizen does not face a cost for doing too little, so first requirement is important.

[2]So that citizen b's utility function is given by $U_b = U_b^M = \gamma \Delta_b + u_b(0) - \frac{c}{2} e_b^2$.

Turning to the equilibrium, citizen a's optimization problem is:

$$\max_{e_a} \ \gamma\left(e_a, \tilde{e}_b\right)\Delta_a + u_a(0) - \frac{c}{2}e_a^2 + \frac{\alpha}{2}\left[\tau\left(\gamma\left(e_a, \tilde{e}_b\right)\Delta_b + u_b(0) - \frac{c}{2}\tilde{e}_b^2\right)\right]$$

The solution to this problem is:

$$e^*\left(\tau\right) = \frac{\Delta_a + \frac{\alpha\tau}{2}\Delta_b}{2c},$$

which by inspection is increasing in the identity parameter τ. Since the propagandist is purely interesting in obtaining the outcome $y = 1$, she wishes to induce the highest level of τ that is affirmed by the citizen.

To understand the affirmation choice by the citizen, i.e., to analyze the demand for different identity conceptions, it is necessary to understand the shape of the value function $V(\tau) \equiv U(e^*(\tau), \tau)$. The derivative of V with respect to τ is equal to:

$$\frac{\alpha}{2}\left[\gamma(e_a^*(\tau), \tilde{e}_b)\Delta_b - \frac{c}{2}\tilde{e}_b^2 + u_b(0)\right]$$

This expression can be positive or negative, depending on parameter values. Thus, sometimes more spiteful and sometimes more altruistic identity conceptions are affirmed.

To investigate this further, note first that in contrast to the specification analyzed in the main text, the value function V is a *convex* function of τ:

$$\frac{\partial^2 V}{\partial \tau^2} = \frac{\alpha\Delta_b^2}{8c} > 0.$$

This means that the function V is generally maximized at a corner solution. To analyze equilibrium attitudes, it is useful to define $\tau_{\min}$ as the value of τ that *minimizes* V. The explicit solution is:

$$\tau_{\min} = \frac{2\left(2cM - \Delta_a\right)}{\alpha\Delta_b}, \quad \text{where} \ \ M \equiv \frac{c\tilde{e}_b^2 - 2u_b(0)}{\Delta_b} - \tilde{e}_b.$$

Moreover, note that e_a^* is increasing in τ if $\Delta_b > 0$ and decreasing otherwise. For simplicity, consider the case of $\Delta_b > 0$, so that the propagandist wishes to install the largest value of τ—the most altruistic attitude—that is still affirmed. There are several cases, see Figure 1 for an illustration.

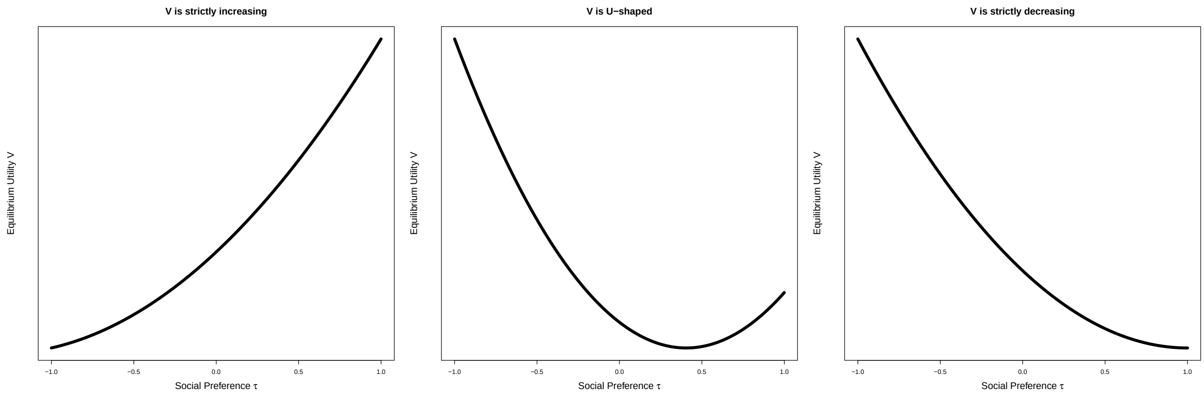

Figure 1: Citizen a's equilibrium utility as a function of the social preference parameter τ. Parameter values: $c = 1$, $u_a(1) = 0.75$, $u_a(0) = 0$, $u_b(1) = 0.2$, $u_b(0) = 0$, $\alpha = 0.5$. Left panel: $\tilde{e}_b = 0.38$; center panel: $\tilde{e}_b = 0.395$, right panel: $\tilde{e}_b = 0.4$.

If $\tau_{\min} < 0$ (left panel), then the citizen's optimal norm τ_{opt} is 1, which is also the propagandist's most preferred value. Hence, the propagandist proposes $\tau_P = 1$, which is affirmed.

If $\tau_{\min} > 1$ (right panel), then V is globally decreasing, and so the citizen's optimal norm τ_{opt} is 0, which is the propagandist's *least* preferred value. The propagandist chooses a value of τ_P that that will not be affirmed, e.g., $\tau_P \geq \tau_E$.

If $\tau_{\min} \in [0, 1]$ (center panel), the citizen's optimal norm τ_{opt} is still 0, but it is possible that a proposed value is preferred to the existing attitude, τ_E. To investigate this formally, not that, because V is a quadratic function of τ, there exists a unique value, $\tau^\dagger$ such that $V(\tau_E) = V(\tau^\dagger)$, i.e., there is an attitude ($\tau^\dagger$) that makes the citizen indifferent between it and the existing attitude τ_E.

The citizen affirms if:

- $\tau_P \notin \left[\tau_E, \tau^\dagger\right]$ if $\tau_E < \tau^\dagger$;
- $\tau_P \notin \left[\tau^\dagger, \tau_E\right]$ if $\tau_E > \tau^\dagger$.

In other words, the citizen affirms if the proposed attitude τ_P is *sufficiently far away* from $\tau_{\min}$. Note that $\tau^\dagger$ can be larger than 1. If so, $\tau_P = 1$ will not be affirmed and the propagandist proposes an attitude τ_P that will not be affirmed (e.g., a value that is close to $\tau_{\min}$ as discussed above). However, if $\tau^\dagger < 1$, the propagandist can propose $\tau_P = 1$ and the citizen will affirm it.

In general, while the analysis is more complicated, it shows that the general approach to identity propaganda can be fruitfully adapted to a different kind of identity conceptualization. One somewhat unsatisfying feature of this approach is that, when effective, equilibrium propaganda is always a corner solution, $\tau_P^* = 1$. Since it is not a function of any parameters, it is of limited use for empirical applications.

3 Identity Propaganda as Salience Messaging

In this section, I consider the implications of allowing the propagandist to choose identity propaganda that increases the salience of identity. In contrast to my treatment of identity norms, I assume that the receiver does not have to affirm the increased salience of identity.[3] However, the propagandist has to pay an exogenous cost to increasing salience.

3.1 Baseline

The propagandist chooses a level of identity salience, $s \in [0, \overline{s}]$. Identity salience α is an increasing and concave function of s. Moreover, changing salience comes at costs $\kappa(s)$, increasing and convex.

The propagandist's optimization problem is:

$$\max_{s \in [0,\overline{s}]} e^*(\alpha(s)) - \kappa(s)$$

Given that $e^*(\alpha(s)) = \frac{\Delta + \alpha(s)\hat{e}}{\alpha(s)+c}$, the first-order condition is:

$$\frac{\hat{e}c - \Delta}{(\alpha + c)^2}\alpha'(s) - \kappa'(s) = 0$$

There is a positive interior solution if and only if $\hat{e} > \frac{\Delta}{c}$. This is intuitive: the propagandist wishes to increase the salience of identity only if that effort level that is prescribed by the identity is higher than the effort level that is materially optimal. Note that if s^* is positive and interior, it is increasing in $\hat{e}$, c and decreasing in Δ. This is an interesting result—the comparative static results are exactly opposite to the baseline analysis of identity norms. This suggests an empirical test could distinguish between identity

[3]Part of the reason is that changing salience is very sensitive to the changing the scale of the function U^I. To see this, observe that optimal effort, $e^* = \frac{\Delta + \alpha\hat{e}}{\alpha + c}$, is increasing in α if and only if $\hat{e} > \frac{\Delta}{c}$. In turn, the value function $V(e^*(\alpha), \alpha)$ is always decreasing in α because U^I is purely a loss function, so $U_I \leq 0$ for all e. If we change U_I to $1 - (e - \hat{e})^2$, which does not change the analysis of identity norms, we have $\frac{\partial V}{\partial \alpha} > 0$.

propaganda as "salience messaging" vs changing identity norms.

3.2 Multiple Social Identities

As before, the propagandist chooses $s \in [0, \bar{s}]$ and maximizes the probability that $y = 1$, i.e., e^*. However, now effort is given by:

$$e^* = \frac{\Delta + \alpha \hat{e}_A + \beta \hat{e}_B}{\alpha + \beta + c}.$$

For the time being, suppose that α is increasing in s and β is decreasing in s, i.e., s increases the salience of being a member in group A but decreases the salience of being a member of group B.

Differentiating the objective function, the first-order condition can be re-arranged to obtain:

$$\alpha'(s)\left[(\beta + c)\hat{e}_A - (\Delta + \beta \hat{e}_B)\right] + \beta'(s)\left[(\alpha + c)\hat{e}_B - (\Delta + \beta \hat{e}_A)\right] - \kappa'(s) = 0$$

A sufficient condition for a positive s^* is that $\hat{e}_A > \frac{\Delta + \beta \hat{e}_B}{\beta + c}$ and $\hat{e}_B < \frac{\Delta + \alpha \hat{e}_A}{\alpha + c}$. Intuitively, if $\hat{e}_A$ is large and $\hat{e}_B$ is small, both conditions are satisfied. To see this more clearly, suppose that $\beta' = -\alpha'$. Then, the first-order condition is:

$$\alpha'(s) \frac{\hat{e}_A - \hat{e}_B}{\alpha + \beta + c} - \kappa'(s) = 0.$$

This shows that $s^* > 0$ if and only if $\hat{e}_A > \hat{e}_B$. Again, this is intuitive—the propagandist only wishes to increases the salience of membership in group A if the effort level that is associated with membership in this group is higher than the effort level that is associated with membership in group B.

4 Applications

4.1 Preference Heterogeneity

Proof of Lemma 3:

Proof. Identical to the proof of Lemma 2. Formally, letting $e^*_{i1} \equiv e^*_i(\hat{e}_{i1})$ and $e^*_{i0} \equiv e^*_i(\hat{e}_0)$, we have that $V(\hat{e}_{i1}) \geq V(\hat{e}_E)$ if:

$$\gamma(e^*_{1i}, e^*_{-i})\Delta_i - \frac{c}{2}(e^*_{i1})^2 - \frac{\alpha}{2}\left(e^*_{i1} - \hat{e}_{i1}\right)^2 \geq \gamma(e^*_{0i}, e^*_{-i})\Delta_i - \frac{c}{2}(e^*_{i0})^2 - \frac{\alpha}{2}\left(e^*_{i0} - \hat{e}_0\right)^2$$

or:

$$\frac{\Delta_i}{2}\left(e^*_{i1} - e^*_{i0}\right) - \frac{c}{2}\left[(e^*_{i1})^2 - (e^*_{i0})^2\right] - \frac{\alpha}{2}\left[(e^*_{i1} - \hat{e}_P)^2 - (e^*_{i0} - \hat{e}_E)^2\right]$$

Replacing Δ_i with $\frac{\Delta_i}{2}$, this is the same expression as expression 1 in Lemma 2. Intuitively, the term e^*_{-i} cancels out when the difference in equilibrium utilities $V(\hat{e}_{i1}) - V(\hat{e}_0)$ is analyzed because the technology γ features additive separable effort choices. □

Proof of Proposition 2:

Proof. Re-arrange inequality 5 to obtain:

$$c\hat{e}_E \geq \Delta_a + \frac{\Delta_b}{2}$$

By inspection, an increase in c or an increase in $\hat{e}_E$ makes it more likely that the inequality is satisfied, i.e., the inequality is satisfied for a larger set of parameter values. As a consequence, the propagandist is more likely to air the more demanding identity norm. □

Generalization Suppose there are n citizens, each choosing e_i and

$$\Pr(y = 1|e_1, ...e_n) = \frac{1}{n}\sum_{i=1}^{n} e_i$$

k of which are extremists with stakes Δ_a and $n-k$ of which are moderates with stakes $\Delta_b < \Delta_a$.

The optimal effort level for citizen i is:

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

Differentiating V_i yields:

$$\frac{dV_i}{d\hat{e}_i} = -\frac{\alpha}{2}\left(e_i^* - \hat{e}_i\right)(-1)$$

which is positive if and only if $\hat{e}_i < \frac{\Delta_i}{nc}$. To give scope to identity norm change for both kinds of citizens, assume that $\hat{e}_E < \frac{\Delta_b}{nc}$.

Since V_i is a polynomial of degree 2 (and hence symmetric around $\hat{e}_{\text{opt}} = \frac{\Delta_i}{nc}$, the identity norm that keeps citizen i indifferent satisfies:

$$\hat{e}_{\text{opt}} - \hat{e}_E = \hat{e}_{i*} - \hat{e}_{\text{opt}}$$

Therefore:

$$\hat{e}_{i*} = 2\frac{\Delta_i}{nc} - \hat{e}_E$$

Define $\hat{e}_* \equiv 2\frac{\Delta_b}{nc} - \hat{e}_E$ and $\hat{e}_{**} \equiv 2\frac{\Delta_a}{nc} - \hat{e}_E$.

For the propagandist, the expected utility of choosing the universally accepted, moderate norm $\hat{e}_*$ is:

$$\frac{ke_a^*(\hat{e}_*) + (n-k)e_b^*(\hat{e}_*)}{n}.$$

By contrast, the expected utility of choosing the divisive, extreme norm $\hat{e}_{**}$ is:

$$\frac{ke_a^*(\hat{e}_{**}) + (n-k)e_b^*(\hat{e}_E)}{n}.$$

Plugging in and re-arranging yields that the latter is optimal if:

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

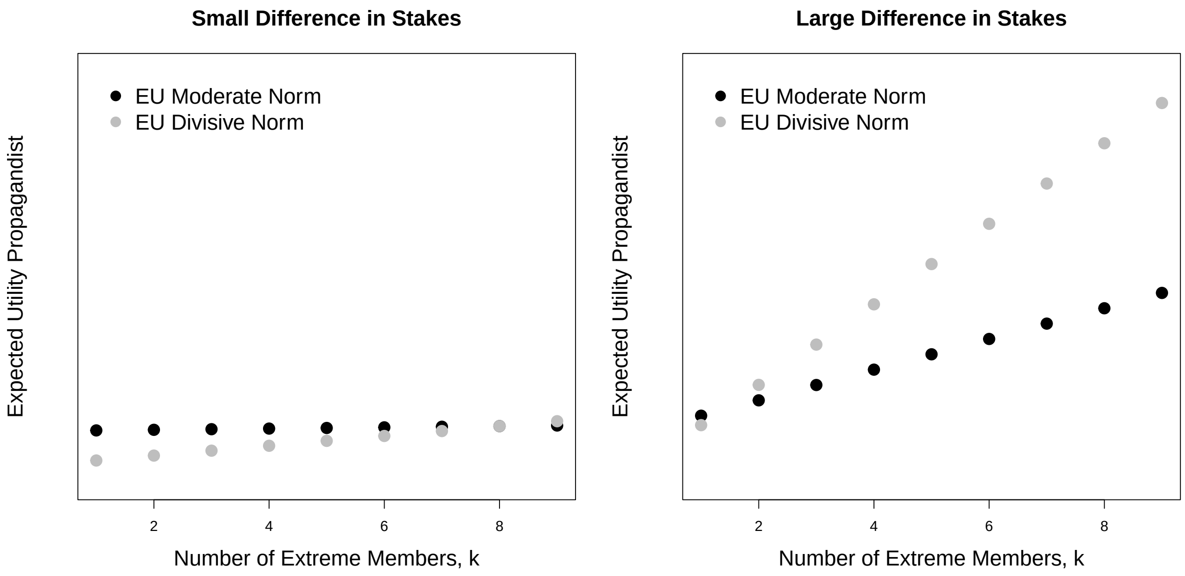

Figure 2: Propagandist's expected utility for choosing moderate vs. divisive norms when citizens are heterogeneous. The x-axis indicates the number of extremists with stakes Δ_a. Parameter values both panels: $\Delta_b = 0.75$, $c = 1.1$, $\alpha = 0.7$, $n = 10$, $\hat{e}_E = 0.05$. Left panel: $\Delta_a = 0.8$; right panel: $\Delta_a = 2$.

If $n = 2$ and $k = 1$, this inequality simplifies to the one in the main text. Note that the right-hand side is decreasing in k: the higher the number of extremists, the more attractive the extremist norm becomes. However, whether or not the divisive norm is optimal also depends on parameter values. To see this, consider Figure 2.

Figure 2 plots the propagandist's expected utility of choosing the moderate, universally accepted norm as well as the expected utility of choosing the divisive, partially accepted norm as a function of the number of extremists, k. When the extremists are only marginally more motivated than the moderated, i.e., the difference in stakes $\Delta_a - \Delta_b$ is small, a substantial number of citizens have to be extreme for the divisive norm to be optimal (left panel). By contrast, when the difference in stakes is larger, the divisive norm can emerge in equilibrium even though the extremists are in the minority (right panel).

4.2 Strategic Interaction among Citizens

Let $\Pr(y = 1|e_a, e_b) \equiv \gamma(e_a, e_b) = \gamma_n(e_a + e_b) + \gamma_s e_a e_b$, with $\gamma_n > 0$. As mentioned in the main text, I assume $\gamma_n > |\gamma_s|$. This ensures that both group members choose interior effort levels.

Equilibrium Mobilization Stage Given commonly known identity norms, citizens solve:

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

The FOC is:

$$\left(\gamma_n + \gamma_s e_{-i}\right)\Delta - ce_i - \alpha\left(e_i - \hat{e}_i\right) = 0$$

This yields the following best response function:

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

So we have the following system of equations:

$$\begin{aligned} e_a(e_b) &= \frac{\left(\gamma_n + \gamma_s e_b\right)\Delta + \alpha\hat{e}_a}{\alpha + c} \\ e_b(e_a) &= \frac{\left(\gamma_n + \gamma_s e_a\right)\Delta + \alpha\hat{e}_b}{\alpha + c} \end{aligned}$$

Solving this system yields:

$$e_a^* = \underbrace{\frac{\gamma_n\Delta(\alpha + c + \gamma_s\Delta)}{(\alpha + c)^2 - \gamma_s^2\Delta^2}}_{\equiv E_0} + \underbrace{\frac{\alpha(\alpha + c)}{(\alpha + c)^2 - \gamma_s^2\Delta^2}}_{\equiv E_1}\hat{e}_a + \underbrace{\frac{\alpha\gamma_s\Delta}{(\alpha + c)^2 - \gamma_s^2\Delta^2}}_{\equiv E_2}\hat{e}_b.$$

Using these definitions, the optimal choices can compactly be written as:

$$e_a^* = E_0 + E_1\hat{e}_a + E_2\hat{e}_b.$$

Similarly:

$$e_b^* = E_0 + E_1\hat{e}_b + E_2\hat{e}_a.$$

Define:

$$V_i(\hat{e}_i) \equiv U_i(e_i^*(\hat{e}_i, \hat{e}_{-i}), e_{-i}^*(\hat{e}_i, \hat{e}_{-i})).$$

The derivative of this function is stated in the main text and details the mechanisms by which a different affirmed identity norm affect i's equilibrium utility.

Now consider when $V_i(\hat{e}_{iP}) \geq V_i(\hat{e}_{iE})$, and note that this influences my own effort choice and the other player's effort choice but not their identity norm.

Keeping $\hat{e}_{-i}$ fixed, i chooses $t_i = 1$ if and only if $V_i(\hat{e}_{iP}, \hat{e}_{-i}) \geq V_i(\hat{e}_{iE}, \hat{e}_{-i})$. Letting $e_i^*(\hat{e}_{iP})$ if $t = 1$ and $e_i^*(\hat{e}_{iE})$ if $t = 0$, we have:

$$\gamma(\hat{e}_P)\Delta - \frac{c}{2}\left(e_{iP}^*\right)^2 - \frac{\alpha}{2}\left[e_{iP}^* - \hat{e}_{iP}\right]^2 \geq \gamma(\hat{e}_E)\Delta - \frac{c}{2}\left(e_{iE}^*\right)^2 - \frac{\alpha}{2}\left[e_{iE}^* - \hat{e}_{iE}\right]^2$$

or:

$$\Delta \underbrace{\left[\gamma(\hat{e}_P) - \gamma(\hat{e}_E)\right]}_{(1)} \geq \frac{c}{2}\underbrace{\left[\left(e_{iP}^* - e_{iE}^*\right)\left(e_{iP}^* + e_{iE}^*\right)\right]}_{(2)} + \frac{\alpha}{2}\underbrace{\left[\left(e_{iP}^* - \hat{e}_{iP}\right)^2 - \left(e_{iE}^* - \hat{e}_{iE}\right)^2\right]}_{(3)}. \quad (3)$$

Write the equilibrium level of effort as:

$$e_{it}^* = E_0 + E_1\hat{e}_{it} + E_2\hat{e}_{-i}\left(t_{-i}\right)$$

where the coefficients E_0, E_1, and E_2 are defined above.

Then, quantity (1) in expression 3 can be re-arranged to obtain:

$$\gamma_n(\hat{e}_{iP}-\hat{e}_{iE})(E_1+E_2)+\gamma_s\left[(\hat{e}_{iP} - \hat{e}_{iE})\left(E_0E_1 + E_0E2 + E_1^2\hat{e}_{-i} + E_1E_2(\hat{e}_{iP} + \hat{e}_{iE}) + E_2^2\hat{e}_{-i}\right)\right]$$

Similarly, quantity (2) in expression 3 can be simplified to obtain:

$$E_1\left(\hat{e}_{iP} - \hat{e}_{iE}\right)\left[2\left(E_0 + E_2\right) + E_1\left(\hat{e}_{iP} + \hat{e}_{iE}\right)\right]$$

Lastly, quantity (3) in expression 3 can be manipulated to obtain:

$$\left(\hat{e}_{iP} - \hat{e}_{iE}\right)\left[2\left(E_0 + E_2\hat{e}_{-i}\right)\left(E_1 - 1\right) + \left(E_1 - 1\right)^2\left(\hat{e}_{iP} + \hat{e}_{iE}\right)\right]$$

Thus, assuming that $\hat{e}_{iP} \geq \hat{e}_{iE}$, we that citizen i affirms if:

$$\Delta(E_1 + E_2)(\gamma_n + E_0\gamma_s) - cE_1(E_0 + E_2) - \alpha(E_1 - 1)E_0 + \hat{e}_{-i}\left[\Delta\gamma_s(E_1^2 + E_2^2) - \alpha(E_1 - 1)E_2\right]$$
$$\geq (\hat{e}_{iP} + \hat{e}_{iE})\left[\frac{c}{2}E_1^2 + \frac{\alpha}{2}(E_1 - 1)^2 - \Delta\gamma_s E_1 E_2\right]$$

We require that $\frac{c}{2}E_1^2 + \frac{\alpha}{2}(E_1 - 1)^2 - \Delta\gamma_s E_1 E_2 > 0$, which requires that γ_s is sufficiently small. Finally, we have that i affirms if and only if:

$$\hat{e}_P \leq T_0 + T_1\hat{e}_{-1}(t_{-i}),$$

where:

$$T_0 \equiv \frac{\Delta(E_1 + E_2)(\gamma_n + E_0\gamma_s) - cE_1(E_0 + E_2) - \alpha(E_1 - 1)E_0}{\frac{c}{2}E_1^2 + \frac{\alpha}{2}(E_1 - 1)^2 - \Delta\gamma_s E_1 E_2} - \hat{e}_{iE}$$
$$T_1 \equiv \frac{\Delta\gamma_s(E_1^2 + E_2^2) + \alpha(1 - E_1)E_2}{\frac{c}{2}E_1^2 + \frac{\alpha}{2}(E_1 - 1)^2 - \Delta\gamma_s E_1 E_2}$$

By inspection, the sign of T_1 is equal to the sign of γ_s.

Following the argument in the main text, there are two critical values of proposed identity content: $\hat{e}'_P = T_0 + T_1\hat{e}_E$ and $\hat{e}''_P = \frac{T_0}{1-T_1}$.

Strategic Complements If $\gamma_s > 0$, then $T_1 > 0$, and so $\hat{e}' < \hat{e}''$. Since $E_1 > 0$ and $E_2 > 0$, both player's effort levels are increasing in their own and in the other citizen's affirmed identity norm. Thus, the propagandist chooses $\hat{e}''_P$ if citizens coordinate on the "both affirm" equilibrium and $\hat{e}'_P$ otherwise.

Strategic Substitutes If $\gamma_s < 0$, then $T_1 < 0$, and so $\hat{e}' > \hat{e}''$. Thus, if $\hat{e}_P \in (\hat{e}_E, \hat{e}''_P)$, player i accept regardless of what the other player is doing. If $\hat{e}_P \in (\hat{e}''_P, \hat{e}'_P)$, they accept only if the other player does not accept. Thus, for the lower interval, there is a symmetric equilibrium in which both players accept. For the upper interval, there are only asymmetric pure strategy equilibria in which exactly one player accepts. Differentiating the propagandist's objective function, it is the case that the probability of getting outcome $y = 1$ is increasing in $\hat{e}_P$ for both intervals; thus, the propagandist chooses between the endpoints. In particular, the propagandist chooses $\hat{e}'_P$ if the following inequality is satisfied:

$$\gamma\left(e_a^*(\hat{e}_P', \hat{e}_P'), e_b^*(\hat{e}_P', \hat{e}_P')\right) \geq \gamma\left(e_a^*(\hat{e}_P'', \hat{e}_E), e_b^*(\hat{e}_P'', \hat{e}_E)\right).$$

Due to the complex interaction between terms, the inequality is difficult to manipulate to generate clear insights. However, it is possible to plot the propagandist's expected utility. In Figure 3, I plot the propagandist's expected utility as a function of $\hat{e}_P$. As the analysis so far indicated, the function peaks twice: at $\hat{e}_P'$ and at $\hat{e}_P''$. Across parameter values, it seems that the propagandist often prefers to air the more moderate identity norm, inducing the symmetric "all affirm" equilibrium with symmetric effort choices.

4.3 Interaction with Material Incentives

Proof of Proposition 4:

Proof. Let $\Psi(r, \nu)$ with $\frac{\partial^2 \Psi}{\partial r \partial \nu} < 0$, i.e., an increase in ν makes repression/cooptation marginally less costly. Naturally, an increase in ν increases the equilibrium amount of repression/cooptation:

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

And since $\hat{e}^* = \frac{2\Delta(r^*(\nu)}{c} - \hat{e}_0$, it follows that $\frac{\partial \hat{e}^*}{\partial \nu} = \frac{2}{c}\frac{\partial \Delta}{\partial r}\frac{\partial r^*}{\partial \nu} > 0$. □

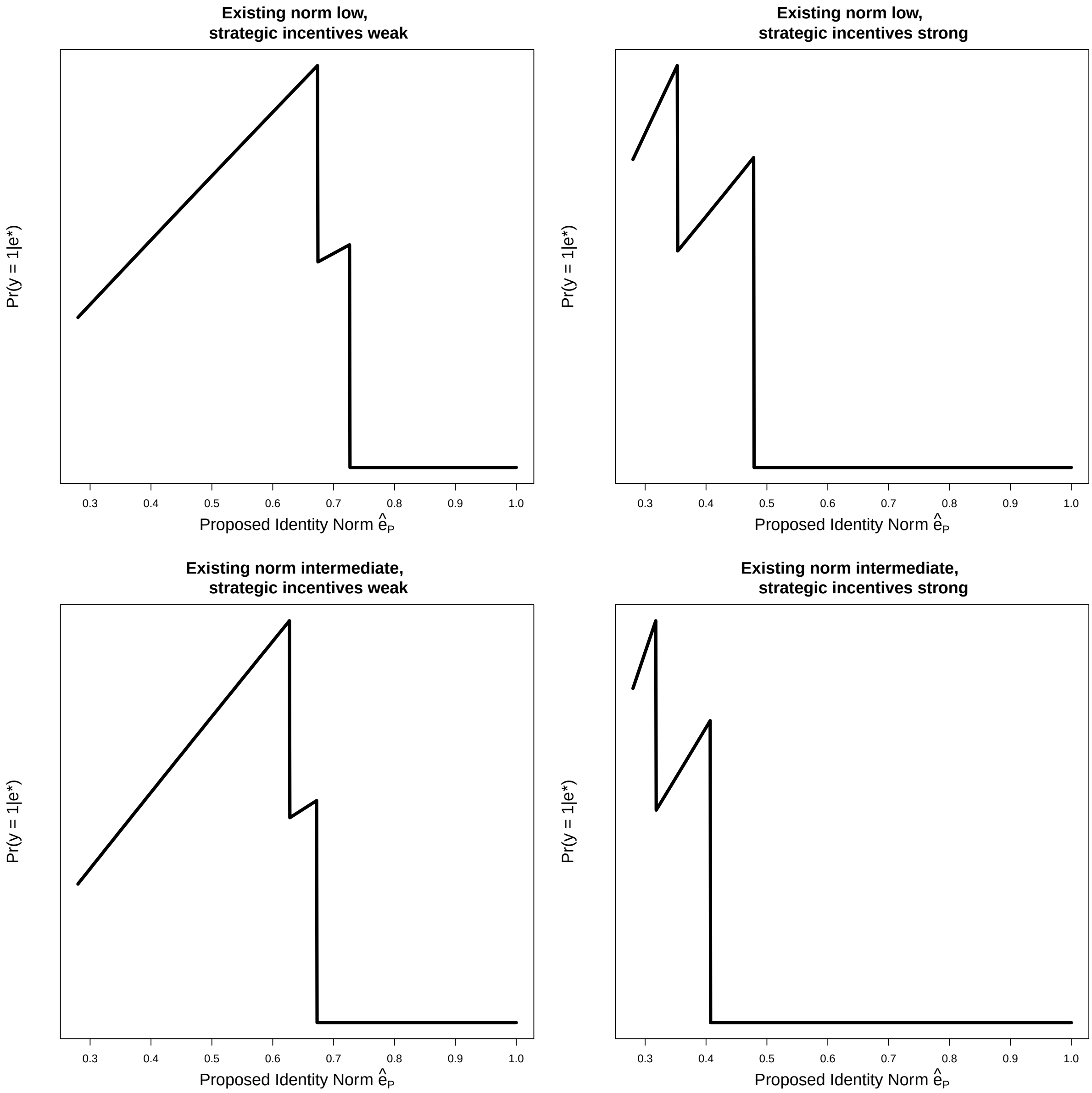

Figure 3: Expected Utility of the Propagandist when effort levels are strategic substitutes. Parameter values for all panels: $\alpha = 0.4$, $c = 1.1$, $\Delta = 1$, $\gamma_n = 0.45$. Rows are $\hat{e}_E = 0.05$ and $\hat{e}_E = 0.1$. Columns are $\gamma_s = -0.1$ and $\gamma_s = -0.4$

5 Additional Analysis

5.1 Competition among Propagandists

In all but the most totalitarian contexts, propagandists cannot expect to be without competition. It is therefore relevant to consider what happens to a propagandist's ability to manipulate identities when there is another propagandist who can also propose a new identity norm. I focus on the case in which the additional leader does not share the propagandist's preferences over political outcomes, i.e., there is a propagandist who wishes to obtain outcome $y = 1$ and a propagandist who wishes to obtain outcome $y = 0$. For concreteness, I refer to the outcome y as regime change, where $y = 1$ means that the regime is overthrown. Hence, the propagandist who wishes to obtain $y = 1$ is the leader of the opposition and denoted by O. The propagandist who wishes to obtain $y = 0$ is the dictator and denoted by D. Hence, $U_O = 1 - y$ and $U_D = y$. Finally, I assume there is only a single citizen, and this citizen chooses the norm proposed by the opposition propagandist when indifferent.

The citizen now chooses between the norm $\hat{e}_O$, the norm $\hat{e}_D$, and the preexisting norm. Recall that the equilibrium utility of the citizen over possible norms is given by the function $V(\hat{e})$ which is uniquely maximized at $\hat{e}_{\text{opt}} = \frac{\Delta}{c}$ and symmetrically decreasing around this point. Consequently, the citizen will choose the norm closest to $\frac{\Delta}{c}$. This leads to complete convergence of identity norms in equilibrium:

Proposition 1. *In the unique equilibrium, propagandists propose identical norms:*

$$\hat{e}_O^* = \hat{e}_D^* = \hat{e}_{opt}$$

The intuition is that the propagandists have to "undercut" each other until this is no longer possible because the citizen does not affirm any other proposed norm. To see this, suppose the opposition propagandist is successful at instilling a more demanding norm as the preexisting identity norm (and the new norm is not equal to the equilibrium utility maximizing norm $\hat{e}_{\text{opt}} = \frac{\Delta}{c}$). Then, the dictator propagandist has an incentive to propose a norm that is lower than the one proposed by the opposition leader and closer to the citizen-preferred norm $\frac{\Delta}{c}$ which the citizen will accept. Consequently, the only stable configuration is both propagandists proposing the norm that maximizes the citizen's equilibrium utility over possible norms. This result is consistent with work emphasizing that media competition can increase citizens' welfare (Gentzkow and Shapiro, 2008). However, it contrasts sharply with intuitions regarding the viability of democracy in diverse societies (e.g., Mansfield and Snyder, 1995): rather than encouraging "outbidding" that results in welfare losses, competition here incentivizes convergences around the citizen's optimal identity norm.

Proof of Proposition 1:

Proof. I first show existence, then uniqueness.

Existence: Given that D (O) chooses $\frac{\Delta}{c}$, no other identity content will be accepted, which implies that O (D) is indifferent among all choices (since no other norm is accepted), so they might as well choose $\frac{\Delta}{c}$.

Uniqueness: I proceed by going through all other cases, showing there is an incentive to deviate for at least one propagandist.

1. Propagandists propose same norm that is different from $\frac{\Delta}{c}$, i.e., $\hat{e}_O = \hat{e}_D \neq \frac{\Delta}{c}$.
    (a) If $\hat{E}_O = \hat{e}_D < \frac{\Delta}{c}$, then O wishes to deviate to a norm in the interval $\left(\hat{E}_O, \frac{2\Delta}{c} - \hat{E}_O\right]$ that is still affirmed but higher than $\hat{E}_O$.
    (b) If $\hat{E}_O = \hat{e}_D > \frac{\Delta}{c}$, then D wishes to deviate to a norm in the interval $\left(\frac{2\Delta}{c} - \hat{e}_D, \hat{e}_D\right)$ that is then affirmed and lower than $\hat{E}_O$.
2. Propagandists propose different norm and one norm is at $\frac{\Delta}{c}$.
    (a) $\hat{e}_D$ is at $\frac{\Delta}{c}$:
        i. $\hat{E}_O < \frac{\Delta}{c} = \hat{e}_D$: D can deviate to a norm in the interval $\left(\hat{E}_O, \frac{\Delta}{c}\right)$ that is still affirmed but lower.
        ii. $\frac{\Delta}{c} = \hat{e}_D < \hat{E}_O$: D can deviate to a norm in the interval $\left(\frac{2\Delta}{c} - \hat{E}_O, \frac{\Delta}{c}\right)$ that is still affirmed but lower.

(b) $\hat{E}_O$ is at $\frac{\Delta}{c}$:

i. $\hat{e}_D < \frac{\Delta}{c} = \hat{E}_O$: O can deviate to a norm in the interval $\left(\frac{\Delta}{c}, 2\hat{e}_D - \frac{\Delta}{c}\right]$ that is still affirmed but higher.

ii. $\frac{\Delta}{c} = \hat{E}_O < \hat{e}_D$: O can deviate to a norm in the interval $\left(\frac{\Delta}{c}, \hat{e}_D\right]$ that is still affirmed but higher.

3. Propagandists propose different norms with $\hat{e}_D < \hat{E}_O$ and neither is at $\frac{\Delta}{c}$:

   (a) Both are to the right: $\frac{\Delta}{c} < \hat{e}_D < \hat{E}_O$: D has an incentive to deviate to a lower norm in the interval $\left(\frac{2\Delta}{c} - \hat{e}_D, \hat{e}_D\right)$ that is still affirmed but lower.

   (b) Both are to the left: $\hat{e}_D < \hat{E}_O < \frac{\Delta}{c}$: O has an incentive to deviate to a higher norm in the interval $\left(\hat{E}_O, \frac{2\Delta}{c} - \hat{E}_O\right]$ that is still affirmed but higher.

   (c) The norms are equidistant: $\hat{e}_D < \frac{\Delta}{c} < \hat{E}_O$ and $|\hat{e}_D - \frac{\Delta}{c}| = |\hat{E}_O - \frac{\Delta}{c}|$: D has an incentive to propose any norm in the interval $\left(\hat{e}_D, \hat{E}_O\right)$ that is then affirmed instead of $\hat{E}_O$.

   (d) $\hat{e}_D$ is closer to $\frac{\Delta}{c}$: O has an incentive to propose any norm in the interval $\left(\hat{e}_D, \frac{2\Delta}{c} - \hat{e}_D\right]$ that is then affirmed and higher than $\hat{e}_D$.

   (e) $\hat{E}_O$ is closer to $\frac{\Delta}{c}$: D has an incentive to propose any norm in the interval $\left(\frac{2\Delta}{c} - \hat{E}_O, \hat{E}_O\right)$ that is then affirmed and lower than $\hat{E}_O$.

4. Propagandists propose different norms with $\hat{E}_O < \hat{e}_D$ and neither is at $\frac{\Delta}{c}$:

   (a) Both norms are to the right of $\frac{\Delta}{c}$: D has an incentive to propose any norm in the interval $\left(\frac{2\Delta}{c} - \hat{E}_O, \hat{E}_O\right)$ that is then affirmed and lower than $\hat{E}_O$.

   (b) Both norms are to the left of $\frac{\Delta}{c}$: O has an incentive to propose any norm in the interval $\left(\hat{e}_D, \frac{2\Delta}{c} - \hat{e}_D\right]$ that is then affirmed and higher than $\hat{e}_D$.

   (c) The norms are equidistant from $\frac{\Delta}{c}$: O has an incentive to choose any norm in the interval $\left(\hat{E}_O, \hat{e}_D\right]$ that is then affirmed and higher than $\hat{E}_O$.

   (d) $\hat{e}_D$ is closer to $\frac{\Delta}{c}$: D has an incentive to choose any norm in the interval $\left[\frac{2\Delta}{c} - \hat{e}_D, \hat{e}_D\right)$ that is still affirmed and lower than $\hat{e}_D$.

   (e) $\hat{E}_O$ is closer to $\frac{\Delta}{c}$: O has an incentive to choose any norm in the interval $\left(\hat{E}_O, \frac{2\Delta}{c} - \hat{E}_O\right]$ that is still affirmed and higher than $\hat{E}_O$.

□

5.2 Multiple Identical Citizens

Suppose there are multiple yet identical citizens (n in total), each of them exerting effort e_i. The technology is given by:

$$\Pr\left(y=1|e_i,e_{-i}\right)=\frac{\sum_{i=1}^{n} e_i}{n}\equiv\gamma(e_i,e_{-i})$$

Note that this implies that all citizens are equally important in bringing about the outcome $y=1$, and effort levels are strategically independent, i.e., $\frac{\partial^2\gamma}{\partial e_i \partial e_{-i}}=0$.

Citizen i's utility function is then:

$$U=yu_i(1)+(1-y)u_i(0)-\frac{c_i}{2}e_i^2-\frac{\alpha_i}{2}\left(e_i-\hat{e}_i\right)^2$$

I assume that ex ante, all citizens are identical: the parameters $u_i(1)$, $u_i(0)$, c_i, α_i, and $\hat{e}_{iE}$ are all the same across citizens. Citizen heterogeneity is explicitly considered in one of the extensions.

Each citizen's optimal choice is:

$$e_i^*=\frac{\frac{\Delta}{n}+\alpha\hat{e}}{\alpha+c}$$

Inspecting the value function $V_i(\hat{e}_i)=U(e_i^*(\hat{e}_i),\hat{e}_i)$ yields the following properties:

- V is increasing in $\hat{e}$ if $\hat{e}<\frac{\Delta}{nc}$, decreasing in $\hat{e}$ if $\hat{e}>\frac{\Delta}{nc}$ and equal to 0 when $\hat{e}=\frac{\Delta}{nc}$.
- The optimal identity norm is $\frac{\Delta}{nc}$.
- If $\hat{e}_E<\frac{\Delta}{nc}$, then i affirms new identity norms if $\hat{e}_P\in\left[\hat{e}_E,2\frac{\Delta}{nc}-\hat{e}_E\right]$

In order to prove these statements, one can proceed in the same fashion as in Lemmas 1, 2, and 3. Given the citizens' affirmation decisions, the propagandist wishes to propose $\frac{2\Delta}{nc}-\hat{e}_E$ if $\hat{e}_E<\frac{\Delta}{nc}$, which is clearly a generalization of the expressions in the main text (if $n=1$, then we have that optimal propaganda equals $\frac{2\Delta}{c}-\hat{e}_E$ whenever $\hat{e}_E<\frac{\Delta}{c}$).

References

Akerlof, George A. and Rachel E. Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics* 115(3):715–752.

Bénabou, Roland and Jean Tirole. 2011. "Identity, Morals and Taboos: Beliefs as Assets." *The Quarterly Journal of Economics* 126(2):805–855.

Chen, Yan and Sherry Xin Li. 2009. "Group identity and social preferences." *American Economic Review* 99(1):431–57.

Fearon, James D. and David D. Laitin. 2000. "Violence and the Social Construction of Ethnic Identity." *International Organization* 54(4):845–877.

Gentzkow, Matthew and Jesse M Shapiro. 2008. "Competition and Truth in the Market for News." *Journal of Economic perspectives* 22(2):133–154.

Mansfield, Edward D. and Jack Snyder. 1995. "Democratization and the Danger of War." *International security* 20(1):5–38.

Shayo, Moses. 2009. "A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution." *American Political Science Review* 103(02):147.

Spenkuch, Jörg L and Philipp Tillmann. 2018. "Elite influence? Religion and the electoral success of the Nazis." *American Journal of Political Science* 62(1):19–36.