Legitimize through Endorsement*

Andrea Gallice† Edoardo Grillo‡

Abstract

The attitude of individuals towards prevailing social norms can change over time. When these changes are not observable, norm abidance can remain high due to social costs. We study how an opinion leader with private information about the societal change can influence norm abidance. We show that the opinion leader can affect abidance when she is neither too ideologically sided in favor of the norm violation, nor too popularity concerned. The opinion leader has a stronger impact on society when social costs are higher, the societal change is more uncertain, and citizens interact more often with like-minded individuals.

JEL Classification: C72, D83.

Keywords: Social norms; societal change; opinion leaders; endorsements; legitimization.

*We thank Sandeep Baliga, Maria Bigoni, Andrea Mattozzi, Moti Michaeli, Daniele Paserman and seminar participants at the 2022 North American (Miami) and European (Milano) Summer Meetings of the Econometric Society, at the 2022 ASSET conference (Crete), at the SIE Annual Meeting (Torino), at the Joint Political Economy and Applied Microeconomics Workshop (Bolzano), at the University of East Anglia, and at Collegio Carlo Alberto for valuable suggestions. This project received funding from the Italian Ministry of University and Research and the European Union-Next Generation EU through the PRIN grant n. 2022WS8AJY "Political Persuasion". Declarations of interest: none.

†Corresponding author. ESOMAS Department, University of Torino, Corso Unione Sovietica 218bis, 10134, Torino, Italy and Collegio Carlo Alberto, Piazza Arbarello 8, 10122, Torino, Italy. *Email*: andrea.gallice@unito.it

‡Department of Economics and Management "Marco Fanno", University of Padova, Via del Santo 22, 35123, Padova, Italy. *Email*: edoardo.grillo@unipd.it

1 Introduction

In 2012, while campaigning to win a second term as President of the United States, Barack Obama publicly declared: "[...] I think same-sex couples should be able to get married." Obama's previous stance on this issue had been more nuanced: he had supported civil unions, while opposing same-sex marriages. The statement got vast media coverage. The net approval toward same-sex marriages in the US increased from +2% in May 2012 to +7% in November 2012, shortly after Obama won reelection.[1]

As the previous example shows, endorsements by prominent public figures can align with, and possibly contribute to, ongoing societal changes. Consider also Oprah Winfrey's endorsement of Obama in the 2008 presidential primaries against Hillary Clinton. According to many political commentators, the endorsement from such a highly influential figure helped convincing many voters that Obama was a serious contendant for the US Presidency.[2] In other circumstances, the lack of an endorsement plays an equally important role. From Michael Jordan's refusal to endorse Harvey Gantt's race to Senate in 1990 North Carolina campaign, to Taylor Swift's lack of support in favor of Hillary Clinton during 2016 presidential campaign,[3] commentators argue that the choice of famous people not to go public with their opinions bring consequences on the electorate.

When does the endorsement (or lack thereof) of an opinion leader influence societal behaviors? When does her endorsement decision have the largest impact?

We address these questions through a model of information transmission enriched with social pressure. In our model, individuals have heterogeneous attitudes towards an established social norm. A shock hits the society and shifts the attitudes of a minority of individuals. For instance, a youth generation becomes more open toward certain civil rights or environmental issues; or the impoverishment of the middle class fosters anti-migrant and xenophobic sentiments among the individuals who fall behind. Individuals in this minority know the extent of the societal change because they experience it first-hand, while the remaining majority is uncertain about it.

Individuals are then randomly matched to play a coordination game in which they must decide whether to abide by the current social norm or to violate it. The interpretation of what constitutes a violation of the norm is broad: it ranges from publicly stating

[1] Source: Gallup polls (`https://news.gallup.com/poll/1651/gay-lesbian-rights.aspx`).

[2] For scholarly work on this, see Garthwaite and Moore (2012).

[3] See `shorturl.at/bnNS4` on Michael Jordan's case and `shorturl.at/ozEY7` on Taylor Swift's one.

a fringe opinion, to taking a controversial action, from breaking a taboo, to violating a commonly accepted routine. Violations encompass both progressive behaviors (e.g., an expansion of civil rights) and regressive ones (e.g., an increase in discriminatory behaviors). Individuals face a coordination problem. If they break the norm, but their match does not, they suffer a social cost. Individuals who conform to a social norm often react with stigma and open hostility against deviant behaviors. The magnitude of this social cost is the *entrenchment of the social norm*.

Before playing the coordination game, individuals can listen to an opinion leader (e.g., a political figure, a religious leader, a celebrity, or a widely-known pundit). The opinion leader holds some imperfect information about the shock that hit the society. We model this assuming that the opinion leader privately observes a binary signal correlated with the shock. A *positive signal* suggests that individuals in the minority are more inclined to violate the norm compared to those in the majority. A *negative signal* suggests that individuals in the minority are less inclined to violate the norm compared to those in the majority. The opinion leader thus has an informational advantage with respect to the majority, and an informational disadvantage with respect to the minority. By virtue of her role, her presence on social media and daily interactions with people, or her preferential access to public opinion polls, the opinion leader is better aware of in-progress societal changes compared to members of the majority. In addition, although some informational spillovers between the minority and the majority are possible, the entrenchment of the social norm limits their relevance and thus preserves the informational advantage of the opinion leader with respect to the majority.

After observing the signal, the opinion leader decides whether to publicly endorse the norm-violating behavior. The opinion leader faces a trade-off. She is ideologically inclined towards the violation of the norm, but the endorsement entails a popularity cost that is proportional to the share of citizens who keep abiding by such norm. The popularity cost can often translate into monetary losses. For instance, when asked about his behavior in the 1990 Senate race, Michael Jordan declared "Republicans buy sneakers, too". In a similar vein, Taylor Swift's popularity among Republican voters at the time likely played a role in her decision not to take side in the 2016 electoral race. Opinion leaders differ in the strength of their ideological motivation against the social norm. We refer to this characteristic as to the opinion leader's *ideological strength*. Opinion leaders with high ideological strength accept large popularity costs to take a stance against the current norm; opinion leaders with low ideological strength do not.

In our baseline model, individuals know the ideological strength of the opinion leader (for instance, thanks to the observation of her past behaviors); we consider the case of uncertain ideological strength in an extension.

Our first result identifies which opinion leaders can hinder or boost societal change through their endorsement decisions. In equilibrium, an opinion leader can modify the behavior of individuals if and only if her ideological strength is neither too high nor too low. An opinion leader with high ideological strength disregards the popularity cost and behaves ideologically: she endorses the norm-violating behavior when she receives a positive signal, but also when she receives a negative signal. In equilibrium, her endorsement provides no information and individuals ignore it. On the contrary, an opinion leader with low ideological strength prefers to avoid the expected popularity cost associated to the endorsement: she refrains from endorsing the violation of the norm when she receives a negative signal, but also when she receives a positive signal. In equilibrium, the lack of endorsement provides no information and individuals ignore it as well. The opinion leader can thus influence societal behavior when she is neither too ideologically sided in favor of the norm violation, nor too concerned about her popularity.

Going back to the introductory example, President Obama's previous, more nuanced, stance on same-sex marriages made him appear non-ideological and gave credibility to his endorsement. At the same time, the popularity cost associated with the statement was not too high: given the increasing support toward same-sex marriages, the endorsement could hardly, by itself, endanger reelection.

We then study how some key features of the society affect *the impact* of the opinion leader. This is the share of individuals who modify their equilibrium behavior when the opinion leader's endorsement (or lack thereof) influences society. If the minority is larger in size, or if there is greater uncertainty concerning the shift in its preferences, the signal of the opinion leader is more informative. Her impact is thus greater. In addition, in societies where the social norm is more entrenched, more individuals abide by the norm despite their private payoffs. The endorsement of the opinion leader has thus a greater potential to modify societal behavior. Also in this case, her impact is greater.

We also investigate how societal features affect *the scope* of the opinion leader. This is the set of opinion leaders who can affect norm abidance. The choice to endorse the norm-violating behavior depends on the comparison between the opinion leader's ideology and the popularity cost she expects to bear. When her signal is more informative about society, the expected popularity cost of an endorsement increases after a negative

signal and decreases after a positive one. Both these changes push the opinion leader to truthfully reveal the signal she received. As a result, if the minority is larger in size, or if there is greater uncertainty concerning the shift in its preferences, the set of opinion leaders who can affect societal behavior widens. Instead, if the entrenchment of the social norm increases, the expected popularity cost of an endorsement increases after both positive and negative signals. Only opinion leaders with relatively stronger ideological motivation can thus affect societal behavior.

Our analysis highlights that opinion leaders are more likely to shape the behavior of societies undergoing potentially deep transformations (e.g., young societies, or societies that experience large migration flows or economic shocks). Furthermore, in societies where a social norm is deep-rooted, successful advocates against it ought to exhibit radical preferences. Our model also shows that the limits in the opinion leader's ability to influence norm-abidance stems from the fact she carries her own agenda and she cannot commit ex-ante to an endorsement strategy.

We then extend the analysis in four directions. First, we introduce homophily: we allow individuals who intrinsically would prefer to violate the norm (or not to violate it) to interact more often with individuals who have a similar attitude. We find that homophily increases the impact of the opinion leader. Second, we show that the insights of our model hold true even if we introduce uncertainty about the opinion leader's ideological strength. Third, we show that a decrease in the informativeness of the signal the opinion leader receives, weakens the impact and scope of her endorsement. Finally, we introduce multiple opinion leaders who independently decide whether to endorse the violation of the norm. The existence of multiple opinion leaders does not affect the credibility of each of them. Yet, multiple endorsement decisions can either reinforce or offset each others. The impact of each opinion leader's endorsement then depends on what all the others do.

1.1 Literature Review

In our model, the opinion leader eases or hinders societal change through her endorsement decision. We thus contribute to the literature that investigates how social norms evolve over time. The bulk of this literature focuses on the long-run evolution of social norms and highlights the role of history (Alesina et al. 2013, Acemoglu and Jackson 2015, Michaeli and Spiro 2017) and institutions (Benabou and Tirole 2011, Acemoglu

and Jackson 2017). In contrast, we study how opinion leaders can shape individuals' behavior in the short-run. In this respect, we are close to Loeper et al. (2014), Carlsson et al. (2016), Bursztyn et al. (2020), Müller and Schwarz (2020), and Grosjean et al. (2021).

Among the papers that take a long-run perspective, Acemoglu and Jackson (2015) is the most related to us. In Acemoglu and Jackson (2015), agents play a coordination game over multiple periods. Some "prominent" individuals with greater visibility can influence the expectations, hence the behavior, of future generations. In Acemoglu and Jackson (2015) individuals differ in their exogenous prominence. In our setting, instead, the preferences of opinion leaders endogenously affect their ability to influence society.

This last feature also distinguishes us from Bursztyn et al. (2020). Like us, Bursztyn et al. (2020) consider a setting in which agents are uncertain about other individuals' preferences and abide by social norms. In their model, individuals update their beliefs about the society based on an exogenous public signal, say a surprising electoral outcome.[4] In this paper, we study the strategic behavior of an opinion leader who provides information through her endorsement decision. This enables us to investigate the interplay between the opinion leader's preferences and societal behavior.

Our paper investigates how the scope and impact of the opinion leader's endorsement decision vary with some key features of the society. This distinguishes our work from Loeper et al. (2014) that study a model in which individuals first observe a random sample of actions taken by biased experts and then decide what to do. Furthermore, in Loeper et al. (2014), experts impact individual choices only if individuals are uncertain about the experts' bias/type. This is due to the joint effect of coordination motives and social learning. In our setting, instead, opinion leaders can affect individual and aggregate behavior even when their types are common knowledge.[5]

We study the credibility of opinion leaders' endorsements. We are thus related to the literature on strategic information transmission. In our setting, the endorsement of the norm-violating behavior entails an ideological benefit and a popularity cost. In this re-

[4]Bursztyn et al. (2020) provide evidence that Trump's victory in the 2016 US presidential election increased individuals' willingness to express xenophobic views and made such opinions more socially acceptable. In a similar vein, Müller and Schwarz (2020) show that Trump's tweets concerning Islam-related topics triggered anti-Muslim hate crimes, whereas Grosjean et al. (2021) find evidence that Trump's rallies boosted racial prejudice against minorities.

[5]In Carlsson et al. (2016), opinion leaders are heterogeneous, but this heterogeneity is in terms of quality rather than ideology/office motivation. Exploiting this heterogeneity, Carlsson et al. (2016) explain why, once in power, some politicians generate consensus on debated issues, while others do not.

spect, we are close to models of information transmission with ideological biases (Cowen and Sutter 1998, Cukierman and Tommasi 1998) and reputational concerns (Morris 2001, Ottaviani and Sørensen 2006a, Ottaviani and Sørensen 2006b). This last feature also links our paper to the literature on pandering (Che et al. 2013, Morelli and Van Weelden 2013, Gratton 2014, and Maskin and Tirole 2019). Within the literature on information transmission, our work is also related to papers that study how experts or politicians can persuade followers (see, among others, Jackson and Tan 2013, Schnakenberg 2015, 2017, Alonso and Câmara 2016, Chan et al. 2019, Gulotty and Luo 2021, Gerardi et al. 2022, Prato and Turner 2022). We contribute to this last literature studying how the ideology of the opinion leader and the norm entrenchment affect her ability to shape societal behavior. The focus on the norm entrenchment also distinguishes us from the literature on media bias (Mullainathan and Shleifer 2005, Baron 2006, Gentzkow and Shapiro 2006, Chiang and Knight 2011, Prat and Strömberg 2013).

Our paper also contributes to the literature on leadership. Hermalin (1998) studies the role of leadership in firms. In Hermalin's model, the leader is the only player that perfectly knows a state of the world. By exerting effort, he can thus induce his coworkers to exert effort as well. We share with Hermalin (1998) (and with most of the literature on leadership) the assumption that the opinion leader has an informational advantage with respect to the majority of citizens, but we differ from it because we assume that the leader has an informational disadvantage with respect to the minority. Chen and Suen (2021) shows that a leader with a strong ideology may exacerbate her radicalism in order to signal the need for a reform and to mobilize citizens in its favor. We differ from Chen and Suen (2021) both in the focus and in the specific modeling assumptions. In particular, we assume that the opinion leader observes a noisy signal of the state. We can then show that both ideology and popularity concerns limit the ability of the opinion leader to affect societal behavior. Block et al. (2021) study the role of leaders and peer pressure in the emergence of consensual or conflictual norms across different social groups.[6] We differ from Block et al. (2021) in that we investigate the credibility of the opinion leader's endorsements as a function of her ideological strength and of her popularity concerns.

[6]Levine et al. (2022) focuses on the role of leaders in conflicts: leaders advise a constituency on the best course of actions in a game of conflict and they can be punished ex-post if their advice leads to an outcome that is worse than what they promised.

The focus on the role for legitimization by an opinion leader in affecting the endogenous evolution of social norms, rather than on the implementation of actual policies that are voted upon, distinguishes us from Chen et al. (2021), that theoretically and empirically study the latter issue in a setting of multiparty political bargaining.

Insofar we study the role of opinion leaders in shaping individual behavior, our work is also related to a recent literature that investigates the market for online endorsements (Fainmesser and Galeotti 2021, Hinnosaar and Hinnosaar 2021, Mitchell 2021). In these models, firms hire influencers to advertise products. The opinion leader in our model does not receive a direct compensation out of her endorsement and there is no third party trying to buy her support.

Finally, in our model, individuals face a coordination problem in the presence of a social norm. Violating this norm entails a social cost. We are thus related to papers that highlight the relevance of social pressure for individual and collective choices (see, for instance, Bernheim 1994, Hopkins and Kornienko 2004, Levy and Razin 2015, Gallice and Grillo 2020, Friedrichsen et al. 2021, and the references therein).

2 The Model

A unit mass of individuals ("he") and an opinion leader ("she") form a society. Individuals belong to one of two groups: a majority M of size $(1-\mu) \in (1/2, 1]$ and a minority m of size μ. The two groups differ in the distribution of individual attitudes toward a prevailing social norm.

Each individual i enjoys an hedonic *private payoff* (i.i.d. within each group) equal to $\theta_i \in \mathbb{R}$ when he violates the norm. The realization of θ_i is private information of individual i. In the majority, the distribution of θ_i is uniform in the interval $[-1, 1]$. In the minority, the distribution of θ_i is uniform in the interval $[\omega - 1, \omega + 1]$. The parameter $\omega \in \mathbb{R}$ measures the difference in the average private payoffs among the two groups; we refer to it as to the *depth of the societal change*. The value of ω is private information of the individuals in the minority. It is uniformly distributed in the interval $[-\psi, \psi]$ from the point of view of individuals in the majority. The parameter $\psi \in \mathbb{R}_{++}$ thus measures the *uncertainty concerning the societal change*. The assumptions concerning uniform distributions provide analytic tractability, but our results immediately extend to other distributions.

Each individual decides independently and simultaneously whether to abide by the norm (action $v_i = 0$), or to violate it (action $v_i = 1$). We refer to individuals choosing $v_i = 0$ as *abiders* and to individuals choosing $v_i = 1$ as *violators*. Individuals are then matched in pairs and get a payoff summarized in Table 1.

i \ j	$v_j = 0$	$v_j = 1$
$v_i = 0$	$0, 0$	$0, \theta_j - \lambda$
$v_i = 1$	$\theta_i - \lambda, 0$	θ_i, θ_j

Table 1: Payoffs from social interaction.

When he abides by the social norm, individual i enjoys a payoff equal to zero. When he violates the norm, instead, he enjoys his private payoff θ_i. The social norm is entrenched: if individual i violates the norm and his match abides by it, the violator suffers a social cost equal to $\lambda \in \mathbb{R}_{++}$, the *norm entrenchment*.

In our baseline model, all matches are equally likely; in Section 5.1, we discuss the case in which individuals are more likely to encounter like-minded individuals. The expected payoff of an individual with private payoff θ_i is thus equal to

$$u(v_i, \overline{v}; \theta_i) = v_i[\theta_i - (1 - \overline{v})\,\lambda], \tag{1}$$

where $\overline{v}$ is the share of violators in the population.

Before individuals choose their actions, an opinion leader can endorse the violation of the social norm (action $e = 1$) or not (action $e = 0$).[7] The opinion leader's endorsement (or lack thereof) becomes common knowledge as soon as it occurs. The opinion leader has private information concerning the depth of the societal change, ω. In particular, the opinion leader observes a private signal $s \in \{0, 1\}$ where:

$$\Pr(s = 0 \mid \omega) = \frac{1}{2} - \frac{\omega}{2\psi} \quad \text{and} \quad \Pr(s = 1 \mid \omega) = \frac{1}{2} + \frac{\omega}{2\psi}.$$

The likelihood of signal $s = 0$ is thus higher than the one of signal $s = 1$ when ω is negative. The opposite is true when ω is positive. We refer to $s = 0$ as to the negative

[7] Given the signal structure discussed below, the focus on a binary action for the opinion leader is without loss of generality.

signal and to $s = 1$ as to the positive signal. A positive (negative) signal suggests that individuals in the minority are on average more (less) inclined to violate the norm than those in the majority.

When the opinion leader chooses not to endorse the violation of the norm, $e = 0$, she gets a payoff equal to 0. When the opinion leader endorses the violation of the norm, $e = 1$, she gets a private payoff equal to $k \in (0, \infty)$, but she also experiences a popularity cost proportional to the share of individuals who keep abiding by the norm. The payoff of the opinion leader is thus equal to:

$$\pi(e; \overline{v}) = e\left[k - (1 - \overline{v})\right]. \tag{2}$$

We define k as the *ideological strength* of the opinion leader. It measures the strength of the opinion leader's private payoff over her popularity concerns. In our baseline model, k is common knowledge. Individuals observe the past record of the opinion leader and identify her attitude towards the social norm. Section 5.2 shows that our insights hold true even when individuals' are uncertain about the opinion leader's ideological strength. Throughout the paper, we maintain that k is positive. If k was negative, the opinion leader would never endorse the violation of the norm because this would generate both a negative ideological payoff and a decrease in popularity. The analysis would then be straightforward.

Assumption 1 below guarantees that the heterogeneity in private payoffs is large enough that some individuals violate the social norm no matter what others do.[8] This implies that the opinion leader's endorsement decision has the potential to modify societal behavior.

Assumption 1. *Some individuals always violate the social norm:* $\max\{\lambda, \psi\} \leqslant 1$.

We solve the game using perfect Bayesian equilibrium. We refer to this solution concept simply as to the equilibrium of the game.

3 Equilibrium Analysis

Individuals in both the majority and the minority optimally follow cutoff strategies. Individual i in group $g \in \{M, m\}$ violates the norm if and only if his private payoff

[8] The assumption also implies that some individuals abide by the norm no matter what others do.

θ_i exceeds a cutoff $\overline{\theta}_g$.[9] Individuals in the minority know the realization of ω and the threshold thus directly depends on ω, $\overline{\theta}_m(\omega)$. Individuals in the majority, instead, do not observe the realization of ω and the threshold depends on ω only through the opinion leader message, $\overline{\theta}_M$.

Let $\overline{v}_M$ and $\overline{v}_m(\omega)$ be the share of violators in the majority and in the minority. The overall share of violators in the society is equal to $\overline{v}(\omega) = (1-\mu)\overline{v}_M + \mu\overline{v}_m(\omega)$.

Proposition 1. *Let $\mathcal{I}_M$ be the information available to the individuals in the majority. The share of violators is equal to:*

$$\overline{v}(\omega) = \overline{v}_b + \frac{\mu}{2-\mu\lambda}\left(\omega + \frac{(1-\mu)\lambda}{2-\lambda}\mathbb{E}\left[\omega \mid \mathcal{I}_M\right]\right),$$

where $\overline{v}_b = \frac{1-\lambda}{2-\lambda}$ is the baseline share of violators, namely the share of violators when there is no social change (i.e., $\mu = 0$).

When $\mu \neq 0$, the share of violators $\overline{v}(\omega)$ differs from the baseline share $\overline{v}_b$ in two respects. First, the shock ω shifts the preferences of the minority. Second, individuals in the majority form an expectation about ω and react to it. This expectation directly impacts the behavior of individuals in the majority, but it also indirectly affects the behavior of individuals in the minority. Although individuals in the minority know the value of ω, payoff function (1) implies that they care about the behavior of the individuals in the majority and they thus react to $\mathbb{E}\left[\omega \mid \mathcal{I}_M\right]$ as well.

The endorsement decision of the opinion leader affects the share of violators through its impact on $\mathbb{E}\left[\omega \mid \mathcal{I}_M\right]$. To understand this effect, we first characterize the benchmark case in which the opinion leader does not exist. In this case, individuals in the majority have no information concerning ω, $\mathbb{E}[\omega \mid \mathcal{I}_M] = 0$.

Remark 1. *When there is no opinion leader, the share of violators is equal to*

$$\overline{v}^{NL}(\omega) = \overline{v}_b + \frac{\mu}{2-\mu\lambda}\omega.$$

When the opinion leader does not exist, the share of violators is higher or lower than the baseline share $\overline{v}_b$ depending on whether the minority is on average more or less inclined to violate the social norm compared to the majority ($\omega > 0$ or $\omega < 0$).

[9] Utility (1) satisfies the single-crossing property in θ_i. The specific tie-breaking rule when $\theta_i = \overline{\theta}_g$ does not affect our analysis.

3.1 Informative Equilibria

Consider now the case in which the opinion leader exists. We first focus on informative equilibria; these are equilibria in which the beliefs of individuals react to the endorsement decision of the opinion leader.

We can summarize the behavior of the opinion leader with an *endorsement strategy*, namely a pair $(\eta(0), \eta(1)) \in [0,1]^2$ where $\eta(s)$ is the probability the opinion leader endorses the violation of the norm (i.e., she chooses $e = 1$) after signal $s \in \{0, 1\}$.

The endorsement strategy is informative if $\eta(0) \neq \eta(1)$. Signal $s = 1$ suggests to the opinion leader that the average private payoff among the minority became larger. Conversely, signal $s = 0$ suggests that the average private payoff among the minority became smaller. If she keeps her behavior constant, the opinion leader then believes that the share of violators is higher after signal $s = 1$ than after signal $s = 0$. We will thus focus on informative endorsement strategies in which $\eta(0) < \eta(1)$: the opinion leader endorses the violation of the norm less often after signal $s = 0$ than after signal $s = 1$.

The endorsement strategy is *fully informative* if $(\eta(0), \eta(1)) = (0, 1)$. In this case, individuals perfectly infer the opinion leader's signal from her endorsement decision. The endorsement strategy is *partially informative* if $0 \leqslant \eta(0) < \eta(1) \leqslant 1$ with at least one of the two weak inequalities being strict. In this case, individuals update their beliefs based on the endorsement decision, but they do not always infer the signal the opinion leader received.

In an informative equilibrium, the expectations of the majority about ω are:

$$\mathbb{E}[\omega \mid e = 0] = -\frac{[\eta(1) - \eta(0)]}{[2 - \eta(1) - \eta(0)]} \cdot \frac{\psi}{3} \quad \text{and} \quad \mathbb{E}[\omega \mid e = 1] = \frac{[\eta(1) - \eta(0)]}{[\eta(1) + \eta(0)]} \cdot \frac{\psi}{3} \tag{3}$$

Proposition 1 then implies that in a fully informative equilibrium the share of violators is greater than $\overline{v}^{NL}(\omega)$ after an endorsement and lower than $\overline{v}^{NL}(\omega)$ after no endorsement.

$$\overline{v}^{FI}(\omega \mid e = 0) = \overline{v}_b + \frac{\mu}{2 - \mu\lambda}\left(\omega - \frac{(1-\mu)\lambda}{2-\lambda} \cdot \frac{\psi}{3}\right), \tag{4}$$

$$\overline{v}^{FI}(\omega \mid e = 1) = \overline{v}_b + \frac{\mu}{2 - \mu\lambda}\left(\omega + \frac{(1-\mu)\lambda}{2-\lambda} \cdot \frac{\psi}{3}\right). \tag{5}$$

The payoff of the opinion leader depends on the share of violators (see equation 2) and this share changes with her endorsement decision. The opinion leader can thus distort her endorsement strategy away from the fully informative one. An opinion leader

with high ideological strength (k large) endorses the violation of the norm even when she receives signal $s = 0$. Her ideological motivation is so strong that she takes a stance against the norm even when the expected popularity cost is high. An opinion leader with strong popularity concerns (k low) does not endorse the violation of the norm even when she receives signal $s = 1$. Because the signal she receives is noisy, the opinion leader incurs an expected popularity cost whenever she endorses the violation. If popularity concerns are strong, she prefers to avoid this cost. A fully informative equilibrium thus exists if and only if the ideological strength of the opinion leader takes intermediate values.

Proposition 2. *A fully informative equilibrium exists if and only if* $k \in \left[\underline{k}, \overline{k}\right]$ *where*

$$\underline{k} = \frac{1}{2-\lambda}\left(1 - \frac{\mu\psi}{3}\right) \quad \text{and} \quad \overline{k} = \frac{1}{2-\lambda}\left(1 + \frac{2(1-\lambda)+\mu\lambda}{2-\mu\lambda} \cdot \frac{\mu\psi}{3}\right).$$

In such equilibrium, the shares of violators are defined by equations (4) *and* (5).

When the ideological strength lies above $\overline{k}$, full information transmission is not possible. Nonetheless, partially informative equilibria exist. In these equilibria, the opinion leader endorses the violation of the norm with certainty after signal $s = 1$, but also, with some positive probability, after signal $s = 0$; that is, $\eta(1) = 1$ and $\eta(0) \in (0, 1)$.[10] The equilibrium share of violators in a partially informative equilibrium is linear in k:

$$\overline{v}^{PI}(\omega \mid e = 0) = k - \frac{\lambda}{2-\lambda} + \frac{\mu}{2-\mu\lambda}\left(\omega - \tfrac{\psi}{3}\right) \tag{6}$$

$$\overline{v}^{PI}(\omega \mid e = 1) = 1 - k + \frac{\mu}{2-\mu\lambda}\left(\omega + \tfrac{\psi}{3}\right). \tag{7}$$

Proposition 3. *A partially informative equilibrium in which* $\eta(0) \in (0, 1)$ *and* $\eta(1) = 1$ *exists if and only if* $k \in \left(\overline{k}, k^{\dagger}\right)$ *where*

$$k^{\dagger} = \frac{1}{2-\lambda}\left(1 + \frac{2-\lambda}{2-\mu\lambda} \cdot \frac{\mu\psi}{3}\right).$$

In such equilibrium, $\eta(0)$ *is increasing in* k *and the shares of violators are defined in equations* (6) *and* (7).

[10] A partially informative equilibrium in which $\eta(0) = 0$ and $\eta(1) \in (0, 1)$ is possible only in the non-generic case in which $k = \underline{k}$ (see the proof of Proposition 3 for details). We will ignore this non-generic case.

When $k = k^\dagger$, equations (6) and (7) are both equal to $\overline{v}^{NL}(\omega)$ and the endorsement decision of the opinion leader does not convey any information.

3.2 Uninformative Equilibria

Uninformative equilibria also exist. In these equilibria, individuals do not update their beliefs based on the opinion leader's endorsement decision; that is, $\mathbb{E}[\omega \mid e = 0] = \mathbb{E}[\omega \mid e = 1] = \mathbb{E}[\omega] = 0$. The share of violators is thus independent of the opinion leader's behavior.

Proposition 4. *An uninformative equilibrium exists if and only if* $k \notin \left[k^U, k^\dagger\right)$, *where*

$$k^U = \frac{1}{2-\lambda}\left(1 - \frac{2-\lambda}{2-\mu\lambda} \cdot \frac{\mu\psi}{3}\right) \in (\underline{k}, \overline{k}).$$

In an uninformative equilibrium, $\overline{v}_2^U(\omega \mid e = 0) = \overline{v}_2^U(\omega \mid e = 1) = \overline{v}_2^{NL}(\omega)$.

Note that uninformative equilibria do not always exist. Unlike in cheap talk models, endorsements carry both an ideological benefit and an expected popularity cost. The opinion leader updates her belief about the popularity cost based on the signal she receives. When she receives signal $s = 1$, she believes that this cost is small. When she receives signal $s = 0$, she believes that this cost is large. When k takes intermediate values, the ideological benefit is above or below the popularity cost depending on the realization of s. The opinion leader thus optimally adjusts her endorsement decision based on the signal and uninformative equilibria do not exist.

For a similar reason, when $k \in [\underline{k}, k^U)$, there exists both a fully informative and an uninformative equilibria. In this range, the ideological strength of the opinion leader (hence, her ideological benefit) is intermediate. She then endorses the violation after signal $s = 1$ if and only if this endorsement leads to a sizable increase in the share of violators.[11] Equilibria with different degrees of informativeness thus exist.

Summary of equilibrium behavior. Figure 1 summarizes the opinion leader's behavior in the equilibria described above. It depicts $\eta(0)$ (solid red line) and $\eta(1)$ (dashed

[11] When k lies above k^U, the opinion leader endorses the violation after receiving signal $s = 1$, independently of the effect this endorsement has on the share of violators. The existence of an informative equilibrium then depends on the incentives of the opinion leader after signal $s = 0$. Converely, when k lies below $\underline{k}$ the opinion leader is unwilling to endorse the violation even if this endorsement has the maximal impact on the share of violators.

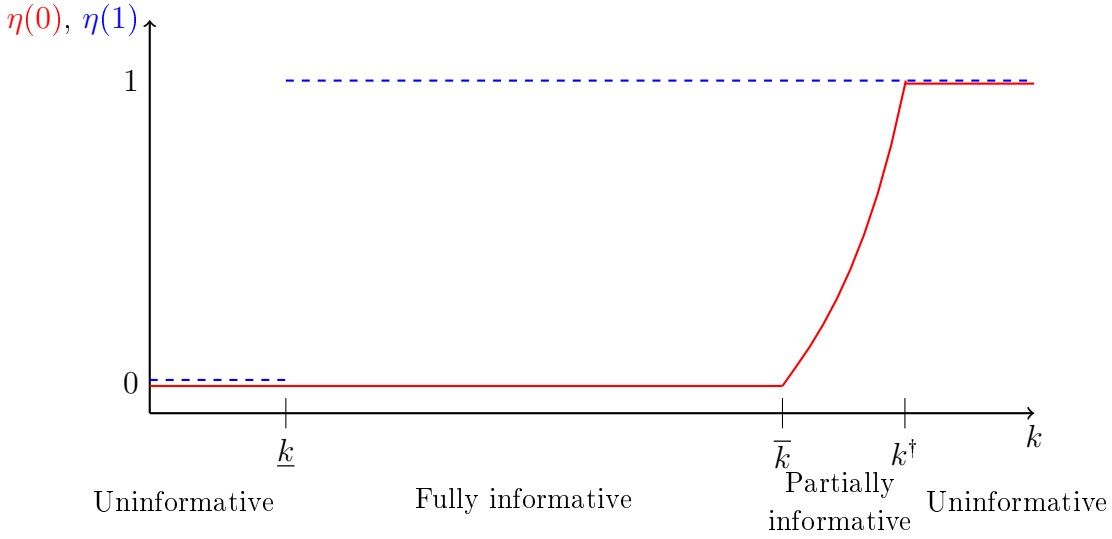

Figure 1: The opinion leader's behavior in the most informative equilibrium.

Notes: The solid red line shows the probability the opinion leader endorses the violation of the norm after signal $s = 0$, $\eta(0)$. The dashed blue line shows the probability the opinion leader endorses the violation of the norm after signal $s = 1$, $\eta(1)$.

blue line) in the most informative equilibrium for different values of k. Outside the ranges identified by Proposition 2 and Proposition 3, only uninformative equilibria exist. When k is below $\underline{k}$, popularity concerns refrain the opinion leader from endorsing the violation of the norm after signal $s = 1$. In this case, $\eta(0) = \eta(1) = 0$. On the contrary, when k exceeds $k^\dagger$, the opinion leader endorses the violation of the norm even after signal $s = 0$. In this case, $\eta(0) = \eta(1) = 1$. Finally, when k takes values in the interval $\left[\underline{k}, k^\dagger\right)$, the opinion leader endorses the violation after signal $s = 1$ and, with a probability that is either equal to zero or positive but smaller than 1, after signal $s = 0$. In this latter case, $\eta(0) \in [0, 1)$ and $\eta(1) = 1$.

4 The Impact of the Opinion Leader

We now discuss the impact of the opinion leader's endorsement decision on norm abidance. This impact is defined as the difference between the share of violators when the opinion leader exists and the same share when she does not exist:

$$\overline{v}(\omega) - \overline{v}^{NL}(\omega) = \frac{\mu}{2 - \mu\lambda} \cdot \frac{(1-\mu)\lambda}{2-\lambda} \, \mathbb{E}\left[\omega \mid \mathcal{I}_M\right]. \tag{8}$$

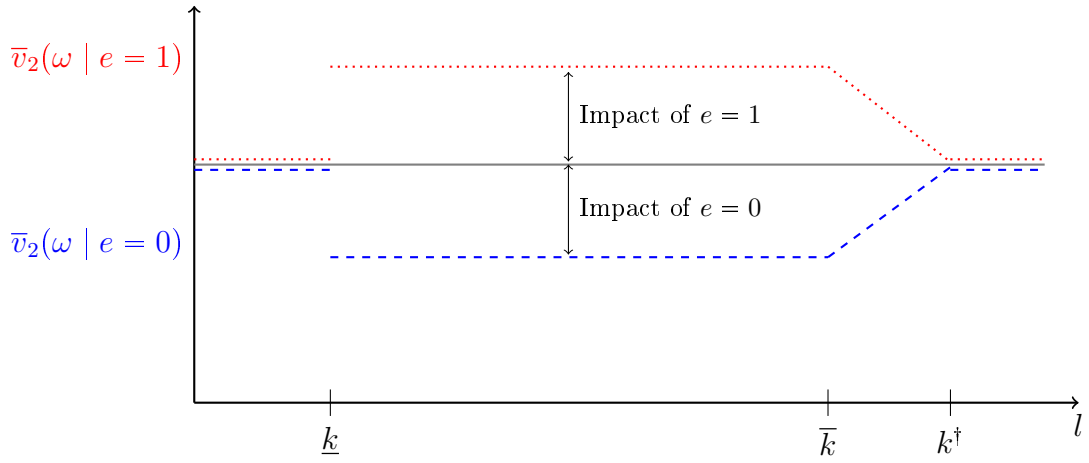

Figure 2: The impact of the opinion leader's endorsement decision.

Notes: The impact of the opinion leader on the share of violators in the most informative equilibrium when $\omega = 0$. The red dotted line shows the share of violators when the opinion leader endorses the violation ($e = 1$). The blue dashed line shows the share of violators when the opinion leader does not endorse the violation ($e = 0$). The gray solid line represents the share of violators when the opinion leader does not exist.

Figure 2 plots the share of violators in the most informative equilibrium when the opinion leader endorses the violation of the social norm (red dotted line), when she does not endorse it (blue dashed line), and when she does not exist (solid gray line). The impact of the opinion leader's endorsement (lack thereof) is equal to the gap between the red dotted line and the gray solid line (the blue dashed line and the gray solid line). These gaps exist only in regions where informative equilibria exist. In a fully informative equilibrium the impact of the opinion leader's endorsement decision is independent of her ideological strength, while in a partially informative equilibrium such impact decreases with ideological strength.

Note that in an informative equilibrium, the share of violators after no endorsement is lower compared to the uninformative equilibrium. This is due to a standard unraveling mechanism. In an informative equilibrium, opinion leaders who receive signal $s = 1$ endorse the violation. The choice not to endorse thus reveals that the opinion leader received signal $s = 0$ and it decreases the share of violators compared to the no-information benchmark.

Proposition 5. *The impact of the opinion leader's endorsement is increasing in μ, λ and ψ. The impact of the opinion leader's lack of endorsement exhibits the opposite comparative statics.*

As the size of the minority μ grows larger, the signal the opinion leader receives conveys information about a larger share of the population. Hence, the expected popularity cost associated with the endorsement goes up after signal $s = 0$, and goes down after signal $s = 1$. This strengthens the incentives of the opinion leader not to endorse the violation of the norm after signal $s = 0$ and to endorse it after signal $s = 1$. The opinion leader's endorsement decision is more credible and thus able to move more people at the margin.

Similarly, when the uncertainty concerning the societal change ψ increases, the signal becomes more informative. The expected cost of endorsing the violation goes up after signal $s = 0$ and it goes down after signal $s = 1$. The opinion leader has thus stronger incentives to match her endorsement decision with the signal she received. Again, her endorsement decision gains credibility.

Finally, when the entrenchment of the norm λ is high, several individuals in the majority refrain from violating the norm fearing social costs. Endorsing the norm violation is more costly for the opinion leader, and thus again more credible. Hence, the opinion leader's decision has more potential to change societal behavior and its impact is thus higher.

Informative equilibria exist if and only if the opinion leader's ideological strength takes intermediate values; that is, if and only if k lies within an upper and a lower bound. We next study how these bounds change with the key parameters in our model.

Proposition 6. *The range* $\left[\underline{k}, \overline{k}\right]$ *for which a fully informative equilibrium exists, widens when* μ *or* ψ *increase, and it shifts to the right when* λ *increases. The range* $\left(\overline{k}, k^{\dagger}\right)$ *for which a partially informative equilibrium exists, shifts to the right when* μ, ψ *or* λ *increase.*

Like in Proposition 5, when the minority represents a sizeable share of the population, or when the uncertainty concerning the societal change is large, the signal of the opinion leader carries more information. As a result, the expected cost of endorsing the violation after signal $s = 0$ ($s = 1$) goes up (goes down). The opinion leader has thus stronger incentives to match her endorsement decision with the signal she received, so that she can more easily establish her credibility. The range $[\underline{k}, \overline{k}]$ unambigously widens.[12] A stronger entrenchment of the social norm implies a higher expected cost from endorsing

[12] We focus on the comparative statics of the range $[\underline{k}, \overline{k}]$. The upper bound on the range of parameters for which a partially informative equilibrium exists, $k^{\dagger}$, reacts to changes in parameters in the same way as $\overline{k}$ does.

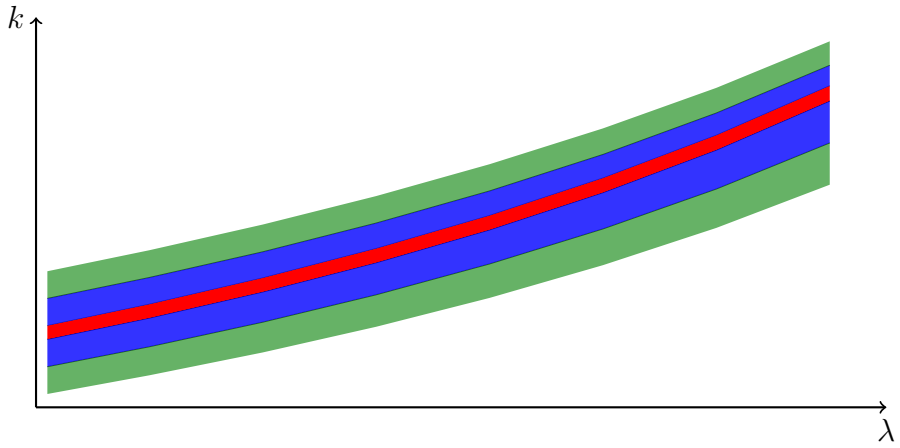

Figure 3: Range of values for which a fully informative equilibrium exists.

Notes: The figure shows the range of k for which a fully informative equilibrium exists as a function of λ when the share of the minority is $\mu = 0.05$ (red), $\mu = 0.25$ (blue) and $\mu = 0.45$ (green).

its violation. Information transmission thus requires higher levels of ideological strength: the range $[\underline{k}, \overline{k}]$ shifts to the right. Figure 3 shows how the range $[\underline{k}, \overline{k}]$ changes with λ and μ.

4.1 The Role of Incentives and Commitment Power

Section 3 shows that the ideology and the popularity concerns of the opinion leader constrain the credibility of her endorsement strategy. To best highlight these effects, suppose the opinion leader is utilitarian and maximizes the sum of the individual utilities. When the depth of the societal change is ω and the majority holds expectation $\mathbb{E}[\omega \mid \mathcal{I}_M]$ over such depth, social welfare is equal to:

$$SW(\omega, \mathbb{E}[\omega \mid \mathcal{I}_M]) = (1-\mu)\int_{\theta_M}^{1} \frac{x - (1-\overline{v}(\omega))\lambda}{2} dx + \mu \int_{\theta_m(\omega)}^{1+\omega} \frac{x - (1-\overline{v}(\omega))\lambda}{2} dx$$

Assume that the opinion leader is uncertain about ω and that she commits to endorsement strategy $(\eta(0), \eta(1))$, knowing that this strategy affects $\mathbb{E}[\omega \mid \mathcal{I}_M]$ as in (3).

Expected social welfare is thus equal to:

$$\mathbb{E}[SW] = \int_{-\psi}^{\psi} \left\{ SW\left(\omega, \frac{\eta(1)-\eta(0)}{\eta(0)+\eta(1)}\frac{\psi}{3}\right)\left[\left(\frac{1}{2}+\frac{\omega}{2\psi}\right)\eta(1)+\left(\frac{1}{2}-\frac{\omega}{2\psi}\right)\eta(0)\right]+ \right.$$
$$\left. SW\left(\omega, \frac{\eta(0)-\eta(1)}{2-\eta(0)-\eta(1)}\frac{\psi}{3}\right)\left[\left(\frac{1}{2}+\frac{\omega}{2\psi}\right)(1-\eta(1))+\left(\frac{1}{2}-\frac{\omega}{2\psi}\right)(1-\eta(0))\right]\right\}\frac{1}{2\psi}d\omega, \tag{9}$$

In this setting, there exists an equilibrium in which the opinion leader chooses a fully informative endorsement strategy, $(\eta(0), \eta(1)) = (0, 1)$. The intuition is straightforward. Two forces drive down the sum of citizens' utilities. First, some citizens abide by the norm despite a positive private payoff. Second, some citizens suffer the social cost because they violate the norm, but their match does not. Full information transmission reduces the likelihood of both these occurrences. When abiders with relatively high private payoffs observe the endorsement, they update upward their expectations concerning the share of violators and they turn into violators. Their expected payoffs thus increase. When violators with relatively low private payoff observe the lack of endorsement, they update downward their expectations concerning the share of violators and they turn into abiders. Their expected payoffs also increase.

Endorsements may lack credibility not only because the opinion leader has her own incentives, but also because she does not have commitment power. We isolate the role of this latter channel considering a setting in which the opinion leader's utility is given by (2), but she can commit ex-ante to an endorsement strategy. In this case, the opinion leader chooses $(\eta(0), \eta(1))$ to maximize

$$\int_{-\psi}^{\psi}\left[\eta(1)\left(k-\left(1-\overline{v}\left(\omega \,\middle|\, \frac{\eta(1)-\eta(0)}{\eta(1)+\eta(0)}\cdot\frac{\psi}{3}\right)\right)\right)\left(\frac{1}{2}+\frac{\omega}{2\psi}\right)\right.$$
$$\left.+\,\eta(0)\left(k-\left(1-\overline{v}\left(\omega \,\middle|\, \frac{\eta(1)-\eta(0)}{\eta(1)+\eta(0)}\cdot\frac{\psi}{3}\right)\right)\right)\left(\frac{1}{2}-\frac{\omega}{2\psi}\right)\right]\frac{d\omega}{2\psi} \tag{10}$$

The optimal endorsement strategy of the opinion leader is as follows. When k is lower than $\underline{k}$, she does not endorse the violation of the norm after either signal. When k is in-between $\underline{k}$ and $\frac{1}{2-\lambda}\left(1+\frac{\mu\psi}{3}\right)$, she endorses the violation when the signal is $s=1$ and she does not endorse it when the signal is $s=0$. Finally, when k is greater than $\frac{1}{2-\lambda}\left(1+\frac{\mu\psi}{3}\right)$, the opinion leader endorses the violation of the norm after both signals.

Because $\frac{1}{2-\lambda}\left(1+\frac{\mu\psi}{3}\right) > k^{\dagger}$, commitment power expands the set of opinion leaders who, despite their strategic incentives, can affect society through their endorsement decisions.

The next proposition summarizes the two benchmarks discussed in this section. Together with the results in Section 3.1, the proposition highlights how the opinion leader's incentives and lack of commitment power affects the credibility of her endorsement.

Proposition 7. *Suppose the opinion leader is utilitarian; that is she maximizes* (9). *She then chooses to always reveal her information,* $(\eta(0),\eta(1)) = (0,1)$.
Suppose the opinion leader's utility is as in (2)*, and that she can commit ex-ante to an endorsement strategy. She then chooses* $(\eta(0),\eta(1)) = (0,0)$ *if* $k \leqslant \underline{k}$, $(\eta(0),\eta(1)) = (0,1)$ *if* $k \in \left(\underline{k}, \frac{1}{2-\lambda}\left(1+\frac{\mu\psi}{3}\right)\right)$, *and* $(\eta(0),\eta(1)) = (1,1)$ *if* $k \geqslant \frac{1}{2-\lambda}\left(1+\frac{\mu\psi}{3}\right)$. *Full information transmission is feasible for a range of ideological strengths wider than in the no-commitment case characterized in Proposition 2.*

5 Extensions

5.1 Homophily

In the baseline model, individuals match with each others with uniform probability. In several social interactions, however, there exists some degree of homophily: individuals with a preference toward or against a social norm attend specific social environments and thus interact more often with individuals who share similar preferences.

To capture this feature, assume that individuals with a positive (negative) private payoff θ_i are more likely to meet other individuals with positive (negative) private payoffs.[13] In particular, the probability an individual with a private payoff $\theta_i \in [-1,1]$ meets an individual with private payoff θ_j is equal to:

$$m\left(\theta_j \mid \theta_i\right) = \begin{cases} \frac{1+h}{(2+h)} & \text{if } \theta_i\theta_j \geqslant 0; \\ \frac{1}{(2+h)} & \text{if } \theta_i\theta_j < 0. \end{cases} \tag{11}$$

[13] We thus model type-dependent homophily. In our model, the behavior of other individuals is uncertain and action-dependent homophily would thus be less sensible (for a discussion of action vs. type-dependent homophily, see Bilancini et al., 2018). Finally, although we model homophily with two groups (those with private payoff greater than zero, and those with private payoff lower than zero), our arguments immediately generalize to settings in which the population is partitioned in any discrete number of groups representing connected intervals of the set of private payoffs.

The parameter $h \in \mathbb{R}_+$ measures the degree of homophily. When $h = 0$, there is no homophily and the model collapses to the baseline one. As h grows, the degree of homophily grows as well. In the limit as $h \to \infty$, individuals with positive (negative) private payoffs interact only with individuals with positive (negative) private payoffs.

Holding constant the expected depth of societal change, $\mathbb{E}[\omega \mid \mathcal{I}_M]$, a higher degree of homophily increases the share of violators.[14] When h increases, individuals interact more often with like-minded individuals. The restraining power of the norm entrenchment on individuals with positive private payoff is thus weaker and the share of violators goes up.

A straightforward adaptation of the proofs of Proposition 2 and Proposition 3 shows that a fully informative and a partially informative equilibrium exist as long as the ideological strength of the opinion leader is neither too high, nor too low. When individuals are matched according to (11), individuals who are pushed into violating the norm by the opinion leader's endorsement meet more often other individuals who are pushed as well. Homophily thus amplifies the impact of the endorsement. Finally, an increase in the degree of homophily also shifts the range for which a fully informative equilibrium exists to the left. When the degree of homophily increases, the share of violators increases too. This lowers the expected cost of the endorsement. Then, both $\underline{k}$ and $\overline{k}$ decrease. Appendix B provides a formal statement of the results discussed in this section.

To sum up, when the society becomes more segregated in echo-chambers and individuals interact more often with people sharing their preferences, the impact of opinion leaders' endorsements grows larger. In addition, the opinion leaders who can influence society are those who are less ideological and more concerned about their popularity. These theoretical predictions are broadly in line with recent political events, namely the rise in ideological polarization paired with the success of political leaders characterized by wavering ideology and populist tendencies (e.g., Donald Trump in the US or Boris Johnson in the UK).

5.2 Uncertainty about the Opinion Leader's Type

In our baseline model, individuals know the ideological strength of the opinion leader; that is, they know k. Often, however, individuals are uncertain about the preferences of the opinion leader. The insights of our paper extend to this case. Suppose that

[14] The comparative statics with respect to the other parameters are as in the baseline model.

individuals believe that the opinion leader's ideological strength is distributed in the interval $[k_\ell, k_h] \subset [0, +\infty)$ according to a continuously differentiable cdf G. Let g be the associated pdf that we assume strictly positive everywhere.

In this setting, we can represent the behavior of the opinion leader with a function $\eta : \{0,1\} \times [k_\ell, k_h] \to [0,1]$, where $\eta(s,k)$ is the probability of an endorsement when the opinion leader has ideological strength k and she received signal s. As in the baseline model, the opinion leader is (weakly) more likely to endorse the violation of the norm after signal $s = 1$ than after signal $s = 0$. The opinion leader is also more likely to endorse the violation of the norm if her ideological strength is higher. The function η is thus increasing in both its arguments.

The uncertainty about k implies that all opinion leaders (but, possibly, a mass of measure zero) play a pure strategy.[15] We can summarize the optimal behavior of the opinion leader with a pair of thresholds (k_0^*, k_1^*). When she receives signal s, the opinion leader endorses the violation of the norm if and only if her ideological strength is greater or equal than k_s^*. The previous discussion implies that $k_0^* \geqslant k_1^*$: if an opinion leader with ideological strength k endorses the violation of the norm after signal $s = 0$, she also does so after signal $s = 1$.

The qualitative features of our baseline model generalizes to the case in which the ideological strength of the opinion leader is uncertain. First, the endorsement decision of the opinion leader is informative if and only if the range of possible ideological strengths, $[k_\ell, k_h]$, is neither shifted too much to the right ($k_\ell > k^\dagger$), nor too much to the left ($k_h < \underline{k}$). If the range of ideological strengths is shifted too much to the right, all opinion leaders endorse the violation of the norm independently of the signal they receive. The endorsement is thus ideologically motivated and lacks credibility. This case arises when, despite a null impact of the endorsement, $k_\ell > k_0^*$. This, in turn, requires $k_\ell > k^\dagger$. If the range of ideological strengths is shifted too much to the left, instead, no opinion leader endorses the violation of the norm fearing the expected popularity cost associated with this action. This case arises when, despite a maximal impact of the endorsement, $k_h < k_1^*$. This, in turn, requires $k_h < \underline{k}$.[16]

[15] Like in Section 3.1, partially informative equilibria still exist. Unlike in Section 3.1, though, the existence of partial information transmission comes from opinion leaders with different ideological strengths playing different pure strategies, rather than from a single opinion leader mixing.

[16] We note that if the opinion leader never endorses the violation of the norm, beliefs after an endorsement are not pinned down. Hence, the expected share of violators after an endorsement, $\mathbb{E}[\overline{v}(\omega) \mid e = 1]$, is thus not pinned down either. However, the incentives to endorse the violation increase with

Second, when the equilibrium is informative, the impact of the endorsement is increasing in the probability that the opinion leader's ideological strength is moderate, namely in the probability that k lies in the interval $[k_0^*, k_1^*]$. Appendix C provides a formal statement of these results.

5.3 Signal's Informativeness

The opinion leader decides whether to endorse the violation of the norm based on a signal concerning the depth of the societal change. This signal represents polls the opinion leader commissioned, or direct contacts she has with various groups of the population. Signals can vary in their informativeness. To capture this heterogeneity, assume that the distribution of the signal conditional on the depth of the societal change is

$$\Pr(s=0 \mid \omega) = (1-\tau)\left(\frac{1}{2} - \frac{\omega}{2\psi}\right) + \frac{\tau}{2} \quad \text{and} \quad \Pr(s=1 \mid \omega) = (1-\tau)\left(\frac{1}{2} + \frac{\omega}{2\psi}\right) + \frac{\tau}{2} \tag{12}$$

As the parameter $\tau \in [0,1]$ increases, the correlation between the signal and the depth of the societal change decreases. The signal structure thus becomes less informative in the spirit of Blackwell (1953). The baseline model in Section 2 corresponds to the case $\tau = 0$. At the other extreme, if $\tau = 1$, the signal is independent of ω: $\Pr[s=0 \mid \omega] = \Pr[s=1 \mid \omega] = 0.5$.

When the informativeness of the signal structure decreases, individuals' expectations react less to the opinion leader's endorsement decision and the impact of her endorsement decreases too.

Less informative signal structures also affect the credibility of the opinion leader: when τ increases, the range of ideological strengths for which there exists a fully informative equilibrium, shrinks. The intuition is simple. The signal now carries less information concerning the underlying societal change. This has two implications. On the one hand, the expected cost of endorsing the violation after $s=1$ increases because the probability of false positive signals (i.e., signal $s=1$ when $\omega < 0$) increases. Only opinion leaders with high ideological strength thus endorse the violation and, as a result, $\underline{k}(\tau)$ increases. On the other hand, the expected cost of endorsing the violation after $s=0$ decreases because the probability of false negative signals (i.e., signal $s=0$ when

$\mathbb{E}[\overline{v}(\omega) \mid e=1]$. Thus, if endorsements are not credible when $\mathbb{E}[\overline{v}(\omega) \mid e=1]$ is maximal, endorsements are not credible even when $\mathbb{E}[\overline{v}(\omega) \mid e=1]$ is less than maximal.

$\omega > 0$) increases. Only endorsements from opinion leaders with strong popularity concerns thus remain credible and, as a result, $\overline{k}(\tau)$ decreases. This latter force is at play also in the partially revealing equilibrium and leads to a reduction in $k^{\dagger}(\tau)$. Appendix D provides a formal statement of these results.

5.4 Multiple Opinion Leaders

Individuals often gather information from multiple sources: they may read several newspapers, listen to multiple pundits on TV, or follow different political leaders on social media. To capture this multiplicity, suppose individuals observe the endorsement decisions of two opinion leaders: $l \in \{1, 2\}$. Let k_l denote the ideological strength of opinion leader l.

Each opinion leader privately receives an independent signal about the depth of the societal change. The signal generating technology is the same for the two opinion leaders and it is the same as in Section 2. Each opinion leader then decides independently and simultaneously whether to endorse the violation of the norm or not. The share of violators is still equal to the expression defined in Proposition 1, but the information set $\mathcal{I}_M$ now includes the endorsement decisions of both opinion leaders.

When multiple opinion leaders exist, the bounds for information transmission characterized in Proposition 2 and Proposition 3 still apply to each opinion leader separately. Opinion leaders move independently and their payoffs (conditional on endorsing the violation) are linear in the share of violators. The law of iterated expectations thus implies that the expected payoff of each opinion leader is identical to the one in our baseline model. In particular, if $k_l \in [\underline{k}, \overline{k}]$ a fully informative equilibrium exists, while if $k_l \in (\overline{k}, k^{\dagger})$ a partially informative equilibrium exists.

Although the existence of multiple opinion leaders does not affect the ability of each of them to convey information, it affects their impact on society. The impact of an opinion leader's endorsement decision now also depends on what the other opinion leaders do. For example, suppose the two opinion leaders play a fully informative strategy. If both opinion leaders endorse the violation or if they both do not, their joint impact is larger (in either direction) than in the model with one opinion leader only. On the contrary, when one opinion leader endorses the violation and the other does not, the overall impact is null and the share of violators is the one in Remark 1.[17]

[17] Compared to Remark 1, individuals' posterior beliefs are now less uncertain; that is, the variance of posterior belief is lower. Because of risk neutrality, this lower variance is irrelevant.

Appendix E provides a more thorough analysis of the case with two opinion leaders. The same logic generalizes to the case in which more than two opinion leaders exist.

6 Conclusions

Prominent political figures, popular media stars and successful social media influencers often affect the behavior of individuals through their actions, statements and endorsements. These opinion leaders modify societal behavior for better or for worse.

In this paper, we study when and to what extent an opinion leader can ease or hinder societal change by endorsing the violation of an established social norm. We build a model in which individuals with heterogeneous propensities to violate the norm are randomly matched and suffer a social cost if they choose to break the norm, while their match does not. A random shock modifies the attitude toward the social norm among a group of individuals in the society. The majority of the society does not know the extent of this shock and may keep abiding by the norm due to social costs. An opinion leader who opposes the norm is partially informed about the shock. Her endorsement of the norm-violating behavior can inform individuals about the extent of the shock and turn some abiders into violators.

We show that the opinion leader's endorsement (or lack thereof) can shape societal behavior when she is neither too ideologically sided against the current norm, nor too popularity concerned. We also show that the impact of the opinion leader's endorsement is larger in societies where the shock to societal preferences is more uncertain and affects a larger share of the population, and in societies where the entrenchment of the norm is higher. Furthermore, the impact of an endorsement is higher in societies where individuals are more likely to interact with other individuals who share their preferences toward the social norm.

Our work highlights how the strategic incentives of opinion leaders affect their ability to ease or hinder societal change. It thus contributes to the current debate on the role and scope of prominent figures in shaping societies.

Appendix

A Proofs

Proof of Proposition 1

Individuals' cutoff strategies are identified by a state-independent threshold for the majority, $\overline{\theta}_M$, and a state-dependent threshold for the minority, $\overline{\theta}_m(\omega)$. Individuals above these threshold are violators, while individuals below them are abiders. The threshold in the minority solves:

$$\overline{\theta}_m(\omega) = \left(1 - (1-\mu)\int_{\overline{\theta}_M}^{1} \frac{dx}{2} - \mu \int_{\overline{\theta}_m(\omega)}^{1+\omega} \frac{dx}{2}\right)\lambda.$$

Solving and rearranging, we get:

$$\overline{\theta}_m(\omega) = \frac{\lambda}{2-\mu\lambda}\left(1 + (1-\mu)\overline{\theta}_M - \mu\omega\right). \tag{A-1}$$

The share of violators in the minority is thus state-dependent and equal to:

$$\overline{v}_m(\omega) = \int_{\overline{\theta}_m(\omega)}^{1+\omega} \frac{dx}{2} = \frac{1}{2} - \frac{\lambda}{2(2-\mu\lambda)}\left(1 + (1-\mu)\overline{\theta}_M\right) + \frac{\omega}{2-\mu\lambda}.$$

Now consider the threshold in the majority; it satisfies the following equation:

$$\overline{\theta}_M = \left(1 - (1-\mu)\int_{\overline{\theta}_M}^{1} \frac{dx}{2} - \mu \int_{-\psi}^{\psi}\left(\int_{\overline{\theta}_m(\omega)}^{1+\omega} \frac{dx}{2}\right) f(\omega \mid \mathcal{I}_M)d\omega\right)\lambda, \tag{A-2}$$

where $f(\omega \mid \mathcal{I}_M)$ is the pdf representing the posterior beliefs about ω by the individuals in the majority when their information set is $\mathcal{I}_M$. Note that:

$$\begin{aligned}\int_{-\psi}^{\psi}\left(\int_{\overline{\theta}_m(\omega)}^{1+\omega} \frac{dx}{2}\right) f(\omega \mid \mathcal{I}_M)d\omega &= \int_{-\psi}^{\psi} \overline{v}(\omega) f(\omega \mid \mathcal{I}_M)d\omega = \mathbb{E}\left[\overline{v}_m(\omega) \mid \mathcal{I}_M\right] = \\ &= \frac{1}{2} - \frac{\lambda}{2(2-\mu\lambda)}\left(1 + (1-\mu)\overline{\theta}_M\right) + \frac{\mathbb{E}[\omega \mid \mathcal{I}_M]}{2-\mu\lambda},\end{aligned}$$

where the last inequality follows from replacing for $\overline{v}_m(\omega)$. If we substitute this expression into equation (A-2) and rearrange, we get:

$$\overline{\theta}_M = \frac{\lambda}{2-\lambda} - \frac{\mu\lambda}{2-\lambda}\mathbb{E}\left[\omega \mid \mathcal{I}_M\right]. \tag{A-3}$$

Plugging equation (A-3) into equation (A-1), we get the cutoff in the minority:

$$\overline{\theta}_m(\omega) = \frac{\lambda}{2-\lambda} - \frac{\mu\lambda}{2-\mu\lambda}\left(\omega + \frac{\lambda(1-\mu)}{2-\lambda}\mathbb{E}\left[\omega \mid \mathcal{I}_M\right]\right). \tag{A-4}$$

The shares of violators in the two groups are thus equal to:

$$\begin{aligned}
\overline{v}_M &= \int_{\overline{\theta}_M}^{1} \frac{dx}{2} = \frac{1-\lambda}{2-\lambda} + \frac{\mu\lambda}{2(2-\lambda)}\mathbb{E}\left[\omega \mid \mathcal{I}_M\right], \\
\overline{v}_m(\omega) &= \int_{\overline{\theta}_m(\omega)}^{1+\omega} \frac{dx}{2} = \frac{1-\lambda}{2-\lambda} + \frac{1}{2-\mu\lambda}\left(\omega + \frac{\lambda^2\mu(1-\mu)}{2(2-\lambda)}\mathbb{E}\left[\omega \mid \mathcal{I}_M\right]\right).
\end{aligned}$$

The overall share of violators follows from taking the weighted sum of these two expression with weights $1-\mu$ and μ. □

Proof of Remark 1

When the opinion leader does not exist, individuals in the majority receive no information concerning ω. Hence, $\mathbb{E}\left[\omega \mid \mathcal{I}_M\right] = 0$. The shares of violators in the two groups become $\overline{v}_M = \frac{1-\lambda}{2-\lambda}$ and $\overline{v}_m(\omega) = \frac{1-\lambda}{2-\lambda} + \frac{1}{2-\mu\lambda}\omega$. The expression for $\overline{v}^{NL}(\omega)$ follows from taking the weighted sum of these two quantities, with $\overline{v}_b = \frac{1-\lambda}{2-\lambda}$ denoting the baseline share of violators when there is no social change (i.e., $\mu = 0$). □

Proof of Proposition 2

When the endorsement strategy is $(\eta(0), \eta(1)) = (0, 1)$, Bayes rule implies that the individuals' posterior beliefs about ω are equal to $f(\omega \mid e = 0) = \frac{1}{2\psi} - \frac{\omega}{2\psi^2}$ and $f(\omega \mid e = 1) = \frac{1}{2\psi} + \frac{\omega}{2\psi^2}$. This implies that the expected values of ω are equal to $\mathbb{E}[\omega \mid e = 0] = -\frac{\psi}{3}$ and $\mathbb{E}[\omega \mid e = 1] = \frac{\psi}{3}$. The share of violators in the society is thus given by equations (4) and (5) in the main text.

In a fully informative equilibrium, the opinion leader endorses the violation of the norm after signal $s = 1$ (first inequality below), and refrain from doing so after signal

$s = 0$ (second inequality):

$$k - \left(1 - \mathbb{E}\left[\overline{v}_b + \frac{\mu}{2-\mu\lambda}\left(\omega + \frac{(1-\mu)\lambda}{2-\lambda}\cdot\frac{\psi}{3}\right) \mid s = 1\right]\right) \geqslant 0$$
$$0 \geqslant k - \left(1 - \mathbb{E}\left[\overline{v}_b + \frac{\mu}{2-\mu\lambda}\left(\omega + \frac{(1-\mu)\lambda}{2-\lambda}\cdot\frac{\psi}{3}\right) \mid s = 0\right]\right)$$

Substituting for $\mathbb{E}[\omega \mid s = 0] = -\frac{\psi}{3}$ and $\mathbb{E}[\omega \mid s = 1] = \frac{\psi}{3}$, we can rewrite the two credibility constraints as:

$$k \geqslant \underline{k} = \frac{1}{2-\lambda}\left(1 - \frac{\mu\psi}{3}\right) \tag{A-5}$$
$$k \leqslant \overline{k} = \frac{1}{2-\lambda}\left(1 + \frac{2(1-\lambda)+\mu\lambda}{2-\mu\lambda}\cdot\frac{\mu\psi}{3}\right) \tag{A-6}$$

These two inequalities define the range $[\underline{k}, \overline{k}]$ in the statement of the proposition. Given that $\psi < 1$, $\underline{k}$ is bounded above zero and $\underline{k} < \overline{k}$. Furthermore, $\overline{k}$ is also bounded above.

Finally, suppose that $k \in [\underline{k}, \overline{k}]$. It is immediate to see that $(\eta(0), \eta(1)) = (0, 1)$ is optimal given the response of individuals specified by equations (4) and (5). Moreover, the cutoff strategies specified in the proof of Proposition 1 are optimal for all individuals when $(\eta(0), \eta(1)) = (0, 1)$. □

Proof of Proposition 3

Suppose there exists an equilibrium in which the opinion leader adopts a partially informative strategy $(\eta(0), \eta(1)) \neq (0, 1)$ with $\eta(0) < \eta(1)$. The expected value of ω conditional on the endorsement decision b would be:

$$\mathbb{E}[\omega \mid e = 0] = -\frac{\eta(1) - \eta(0)}{2 - \eta(1) - \eta(0)}\cdot\frac{\psi}{3} \qquad \text{and} \qquad \mathbb{E}[\omega \mid e = 1] = \frac{\eta(1) - \eta(0)}{\eta(1) + \eta(0)}\cdot\frac{\psi}{3}.$$

In equilibrium, the opinion leader must be willing to choose $e = 1$ with probability $\eta(1)$ after signal $s = 1$, and to choose $e = 1$ with probability $\eta(0)$ after signal $s = 0$. Consider the case in which $\eta(0) \in (0, 1)$ and $\eta(1) = 1$. The opinion leader must endorse the violation after $s = 1$ and be indifferent between endorsing or not after signal $s = 0$. The

following two conditions must hold:

$$k \geqslant \frac{1}{2-\lambda} - \frac{\mu}{2-\mu\lambda}\left(\frac{(1-\mu)\lambda}{2-\lambda}\cdot\frac{1-\eta(0)}{1+\eta(0)}+1\right)\cdot\frac{\psi}{3}$$
$$k = \frac{1}{2-\lambda} - \frac{\mu}{2-\mu\lambda}\left(\frac{(1-\mu)\lambda}{2-\lambda}\cdot\frac{1-\eta(0)}{1+\eta(0)}-1\right)\cdot\frac{\psi}{3}.$$

From the equality, we get:

$$\eta(0) = 1 - 2\cdot\frac{\frac{1}{\lambda(1-\mu)}\left(1-\frac{3(2-\mu\lambda)}{\mu\psi}\left(k-\frac{1}{2-\lambda}\right)\right)(2-\lambda)}{1+\frac{1}{\lambda(1-\mu)}\left(1-\frac{3(2-\mu\lambda)}{\mu\psi}\left(k-\frac{1}{2-\lambda}\right)\right)(2-\lambda)}. \tag{A-7}$$

The right-hand side of equation (A-7) is increasing in k. Furthermore, $\eta(0) > 0$ if

$$k > \overline{k} = \frac{1}{2-\lambda}\left(1+\frac{2(1-\lambda)+\mu\lambda}{2-\mu\lambda}\cdot\frac{\mu\psi}{3}\right)$$

and $\eta(0) < 1$ if

$$k < k^{\dagger} = \frac{1}{2-\lambda}\left(1+\frac{2-\lambda}{2-\mu\lambda}\cdot\frac{\mu\psi}{3}\right).$$

This proves the existence of a partially informative equilibrium in which $\eta(0) \in (0,1)$ and $\eta(1) = 1$ for any $k \in (\overline{k}, k^{\dagger})$. If we substitute the $\eta(0)$ and $\eta(1)$ we just obtained in the overall share of violators, we obtain equations (6) and (7).

Now, suppose there exists an equilibrium in which $\eta(0) = 0$ and $\eta(1) \in (0,1)$. In this case, we would need:

$$k = \frac{1}{2-\lambda} - \frac{\mu}{2-\mu\lambda}\left(\frac{(1-\mu)\lambda}{2-\lambda}+1\right)\cdot\frac{\psi}{3}$$
$$k \leqslant \frac{1}{2-\lambda} - \frac{\mu}{2-\mu\lambda}\left(\frac{(1-\mu)\lambda}{2-\lambda}-1\right)\cdot\frac{\psi}{3}$$

Hence, a partially informative equilibrium exists if and only if k is non-generic and equal to $\underline{k} = \frac{1}{2-\lambda}\left(1-\frac{\mu\psi}{3}\right)$. □

Proof of Proposition 4

In an uninformative equilibrium, the expectations of individuals do not react to the endorsement decision of the opinion leader: $\mathbb{E}[\omega \mid e=0] = \mathbb{E}[\omega \mid e=1] = \mathbb{E}[\omega] = 0$.

This happen when, in equilibrium, the opinion leader does not modify her behavior based on the signal she receives. We can thus have two possible scenarios.

First, the opinion leader may endorse the violation of the norm no matter which signal she received. Optimality requires $\eta(1) \geqslant \eta(0)$. This first scenario arises when the opinion leader endorses the violation of the norm after signal $s = 0$; namely when:

$$k \geqslant \frac{1}{2-\lambda}\left(1 + \frac{2-\lambda}{2-\mu\lambda} \cdot \frac{\mu\psi}{3}\right) = k^{\dagger}.$$

Second, the opinion leader may not endorse the violation of the norm no matter which signal she received. Because optimality requires $\eta(0) < \eta(1)$, this second scenario arises when the opinion leader does not endorse the violation after signal $s = 1$, namely when

$$k \leqslant \frac{1}{2-\lambda}\left[1 - \frac{2-\lambda}{2-\mu\lambda} \cdot \frac{\mu\psi}{3}\right] := k^{U} \in (\underline{k}, \overline{k}).$$

Proof of Proposition 5

In a fully informative equilibrium, we have that:

$$\overline{v}(\omega \mid e = 0) - \overline{v}^{NL}(\omega) = -\frac{\mu}{2-\mu\lambda} \cdot \frac{(1-\mu)\lambda\psi}{3(2-\lambda)}$$
$$\overline{v}(\omega \mid e = 1) - \overline{v}^{NL}(\omega) = \frac{\mu}{2-\mu\lambda} \cdot \frac{(1-\mu)\lambda\psi}{3(2-\lambda)}$$

Hence, in a fully informative equilibrium, the impact of the opinion leader's endorsement, $e = 1$, on the share of violators is increasing in ψ and λ. The impact of a lack of endorsement is reversed. Finally, the derivative of the opinion leader's impact with respect to μ after an endorsement is equal to $\frac{\lambda\psi(2-4\mu+\mu^2\lambda)}{3(2-\mu\lambda)^2(2-\lambda)}$. This expression is always positive because $\mu < \frac{1}{2}$. The impact of the lack of an endorsement is symmetric with opposite sign.

In a partially informative equilibrium, instead, the impact of the opinion leader is asymmetric depending on whether she endorses the violation or not. If the opinion leader endorses the violation, her impact is

$$1 - k - \frac{1-\lambda}{2-\lambda} + \frac{\mu}{2-\mu\lambda} \cdot \frac{\psi}{3}.$$

By Assumption 1, this expression is increasing in μ, λ and ψ, while it is decreasing in k. If the opinion leader does not endorse the violation, her impact is

$$k - \frac{1}{2-\lambda} - \frac{\mu}{2-\mu\lambda} \cdot \frac{\psi}{3}.$$

By Assumption 1, this expression is decreasing in μ, λ and ψ, and increasing in k. □

Proof of Proposition 6

First, consider $\underline{k} = \frac{1}{2-\lambda}\left(1 - \frac{\mu\psi}{3}\right)$. It is immediate to verify that this bound is decreasing in μ and ψ, while it is increasing in λ.

Now consider $\overline{k} = \frac{1}{2-\lambda}\left(1 + \frac{2(1-\lambda)+\mu\lambda}{2-\mu\lambda} \cdot \frac{\mu\psi}{3}\right)$. This bound is increasing in ψ. Consider the derivative of $\overline{k}$ with respect to μ:

$$\frac{\partial \overline{k}}{\partial \lambda} = \frac{4(3 - 3\mu\lambda - \psi\mu(1-\mu)) + 4\mu^2\psi(1-\mu) + \mu^2\lambda^2(3 + \psi(2-\mu))}{3(2-\lambda)^2(2-\mu\lambda)^2}.$$

Note that $4\left(3 - 3\mu\lambda - \psi\mu(1-\mu)\right) > 12(1 - \lambda/2 - \psi/4) > 0$, where the first inequality follows from $\mu < 1/2$ and the second one from Assumption 1. We conclude that $\overline{k}$ is increasing in λ. The derivative with respect to μ is equal to:

$$\frac{\partial \overline{k}}{\partial \mu} = \frac{4 - \mu^2\lambda^2 - 4(1-\mu)\lambda}{3\mu^2\lambda^2(2-\lambda) + 12(2-\lambda)(1-\mu\lambda)}\psi.$$

This expression is positive since both the numerator and the denominator are positive (see Assumption 1). Thus $\overline{k}$ increases with μ. The results on the range of values of k for which the fully informative equilibrium exists, $[\underline{k}, \overline{k}]$, follow immediately from the previous analysis.

Now consider the partially informative equilibria. It is easy to verify that $k^\dagger = \frac{1}{2-\lambda}\left(1 + \frac{2-\lambda}{2-\mu\lambda} \cdot \frac{\mu\psi}{3}\right)$ is increasing in ψ and μ. $k^\dagger$ is also increasing in λ. Indeed, by Assumption 1 we have

$$\frac{\partial k^\dagger}{\partial \lambda} = \frac{\mu^2(3\lambda + \psi) + 4\mu^2\psi(1-\lambda) + 12(1-\mu\lambda)}{3(2-\lambda)^2(2-\mu\lambda)^2} > 0.$$

The results on the range of values of k for which a partially informative equilibrium exists, $\left(\overline{k}, k^\dagger\right)$, follow from the derivatives computed above. □

Proof of Proposition 7

Consider the expected social welfare $\mathbb{E}[SW]$ in the main text. The vector of first order derivatives with respect to $\eta(0)$ and $\eta(1)$ is equal to:

$$\frac{1}{18}\mu^2\lambda^2\psi^2\frac{1-\mu}{(2-\mu\lambda)^2(\lambda-2)^2}\left(12-4\lambda(1+\mu)+\mu\lambda^2\right)\times \begin{bmatrix} -\frac{[\eta(1)-\eta(0)][3\eta(1)+\eta(0)-2\eta(0)\eta(1)-2\eta(1)^2]}{[(2-\eta(0)-\eta(1))(\eta(0)+\eta(1))]^2} \\ \frac{[\eta(1)-\eta(0)][3\eta(0)+\eta(1)-2\eta(0)\eta(1)-2\eta(0)^2]}{[(2-\eta(0)-\eta(1))(\eta(0)+\eta(1))]^2} \end{bmatrix}$$

The first order necessary conditions are equal to 0 if $\eta(0)=\eta(1)$ or if $\eta(0)=-\eta(1)$. In both cases, the objective function increases if we (weakly) increase $\eta(1)$ and we (weakly) decrease $\eta(0)$. The optimal endorsement strategy is thus equal to $(\eta(0),\eta(1))=(0,1)$.

Consider now an opinion leader with ideological strength equal to k and the ability to commit. The derivatives of (10) with respect to $\eta(0)$ and $\eta(1)$ are respectively: $\frac{(6k+\mu\psi-3k\lambda-3)}{6(2-\lambda)}$ and $\frac{(6k-\mu\psi-3k\lambda-3)}{6(2-\lambda)}$. The values of $\eta(0)$ and $\eta(1)$ are thus corner solutions. The statement of the proposition follows from checking when these derivatives are positive. □

B Homophily: Formal Results

In this section we provide a formal statement (and the related proof) of the results on homophily discussed in Section 5.1.

Proposition B.1. *When social interactions are characterized by a degree of homophily equal to $h\in\mathbb{R}_+$, the share of violators is equal to*

$$\bar{v}(\omega,h)=\bar{v}_b(h)+\frac{2+h}{2}\cdot\frac{\mu}{2+h-(1+h)\mu\lambda}\cdot\left(\omega+\frac{(1-\mu)(1+h)\lambda}{2+h-(1+h)\lambda}\mathbb{E}\left[\omega\mid\mathcal{I}_M\right]\right)$$

where $\bar{v}_b(h)=\frac{2+h}{2}\left(\frac{1-\lambda}{2+h-(1+h)\lambda}\right)$. The share $\bar{v}(\omega,h)$ is increasing in h and so it is the impact of the opinion leader's endorsement. Furthermore, as h increases the range $[\underline{k},\overline{k}]$ for which a fully informative equilibrium exists shifts to the left, while the range $(\overline{k},k^\dagger)$ for which a partially informative equilibrium exists can either shift to the left or widen.

Proof. The same logic used in the proof of Proposition 1 implies that we can then define two cutoff strategies. Individuals in the majority adopt a state-independent cutoff

strategy with thresholds $\overline{\theta}_M(h)$. Individuals in the minority adopt a state-dependent cutoff strategy with threshold $\overline{\theta}_m(\omega, h)$. The two thresholds are given by:

$$\begin{aligned}
\overline{\theta}_M(h) &= \frac{\lambda}{2+h-(1+h)\lambda} - \frac{(1+h)\mu\lambda}{2+h-(1+h)\lambda}\mathbb{E}\left[\omega \mid \mathcal{I}_M\right] \\
\overline{\theta}_m(\omega, h) &= \frac{\lambda}{2+h-(1+h)\lambda} - \\
&\quad - \frac{(1+h)\mu\lambda}{2+h-(1+h)\mu\lambda}\left(\omega + \frac{(1-\mu)(1+h)\lambda}{2+h-(1+h)\lambda}\mathbb{E}\left[\omega \mid \mathcal{I}_M\right]\right)
\end{aligned}$$

The share of violators in the two groups are then given by:

$$\begin{aligned}
\overline{v}_M(h) &= \int_{\bar{\theta}_M(h)}^{1} \frac{dx}{2} = \bar{v}_b(h) + \frac{1}{2}\cdot\frac{\lambda(1+h)\mu}{(2+h-(1+h)\lambda}\mathbb{E}\left[\omega \mid \mathcal{I}_M\right] \\
\overline{v}_m(\omega, h) &= \int_{\bar{\theta}_m(\omega,h)}^{1+\omega} \frac{dx}{2} = \overline{v}_b(h) \\
&\quad + \frac{1}{2+h-(1+h)\mu\lambda}\left(\frac{2+h}{2}\omega + \frac{\mu(1-\mu)(1+h)^2\lambda^2}{2\left[2+h-(1+h)\lambda\right]}\mathbb{E}\left[\omega \mid \mathcal{I}_M\right]\right)
\end{aligned}$$

The overall share of violators is obtained taking the weighted sum of $\overline{v}_M(h)$ and $\overline{v}_m(\omega, h)$:

$$\begin{aligned}
\overline{v}(\omega, h) &= (1-\mu)\,\overline{v}_M(h) + \mu\overline{v}_m(\omega, h) = \\
&\quad \overline{v}_b(h) + \frac{2+h}{2}\cdot\frac{\mu}{2+h-(1+h)\mu\lambda}\cdot\left(\omega + \frac{(1-\mu)(1+h)\lambda}{2+h-(1+h)\lambda}\mathbb{E}\left[\omega \mid \mathcal{I}_M\right]\right)
\end{aligned}$$

This share is increasing in the degree of homophily. To see why, notice that $\overline{v}_b(h)$ is increasing in h. Then observe that the derivatives with respect to h of the terms in ω and in $\mathbb{E}\left[\omega \mid \mathcal{I}_M\right]$ are proportional to ω and $\mathbb{E}\left[\omega \mid \mathcal{I}_M\right]$. Hence, the overall derivative with respect to h is minimized when $\omega = \mathbb{E}\left[\omega \mid \mathcal{I}_M\right] = -\psi$. This minimal value is positive. Hence, $\overline{v}(\omega, h)$ is increasing in h. The same argument also proves that the impact of the opinion leader's endorsement is increasing in h.

If we replicate the steps of the proof of Proposition 2 and we take into account that in a fully informative equilibrium we still have $\mathbb{E}\left[\omega \mid e=0\right] = -\frac{\psi}{3}$ and $\mathbb{E}\left[\omega \mid e=1\right] = \frac{\psi}{3}$, one obtains that a fully informative equilibrium exists if and only if $k \in [\underline{k}(h), \overline{k}(h)]$,

where:

$$\underline{k}(h) = \frac{1}{2\left(2+h\right)-\left(1+h\right)\lambda]}\left(2+h-h\lambda-\frac{\left(2+h\right)\mu\psi}{3}\right)$$
$$\overline{k}(h) = \frac{1}{2\left[\left(2+h\right)-\left(1+h\right)\lambda\right]}\Bigg(2+h-h\lambda+$$
$$+\frac{2+h-\left(1+h\right)\lambda-\left(1+h\right)\left(1-\mu\right)\lambda}{2+h-\left(1+h\right)\mu\lambda}\cdot\frac{\left(2+h\right)\mu\psi}{3}\Bigg).$$

The derivative of $\underline{k}(h)$ with respect to h is given by:

$$\frac{\partial\underline{k}(h)}{\partial h} = -\lambda\frac{3\left(1-\lambda\right)+\mu\psi}{6\left(2+h-\left(1+h\right)\lambda\right)^2} < 0$$

while the derivative of $\overline{k}(h)$ with respect to h is given by:

$$\frac{\partial\overline{k}(h)}{\partial h} = -\lambda\frac{3(1-\lambda)-\mu\psi}{6[2+h-(1+h)\lambda]} - \frac{\mu\lambda\psi}{3}\frac{(1-\mu)[(4(1+h)+h^2)-\mu\lambda^2(1+h)^2]}{[2+h-(1+h)\lambda]^2[2+h-(1+h)\mu\lambda]^2}.$$

This expression is bounded above by

$$-\frac{\mu\lambda\psi}{6[2+h-(1+h)\lambda]^2}\frac{2(2+h)-(1+h)(1+\mu)}{[2+h-(1+h)\mu\lambda]^2}\mu\lambda^2(1+h),$$

which is negative. Hence, both $\underline{k}(h)$ and $\overline{k}(h)$ are decreasing in h.

Now consider partially informative equilibria in which $\eta(1) = 1$ and $\eta(0) \in (0,1)$ The same steps of Proposition 3 imply that the upper bound $k^{\dagger}(h)$ is equal to:

$$k^{\dagger}(h) = \frac{1}{2}\left[1+\frac{\lambda}{2+h-(1+h)\lambda}+\frac{(2+h)\mu\psi}{3[2+h-(1+h)\mu\lambda]}\right].$$

The derivative of this expression with respect to h can be either negative or positive depending on whether $(1-\lambda)/[2+h-(1+h)\lambda]^2$ is greater or lower than $\mu\psi/[2+h-(1+h)\mu\lambda]^2$. □

C Uncertainty about the Opinion Leader's Ideological Strength: Formal Results

In this section we provide a formal statement (and the related proof) of the results about the robustness of our insights to the case in which the ideological strength of the opinion leader is uncertain (see Section 5.2).

Proposition C.1. *Suppose that the ideological strength of the opinion leader is distributed in the interval $[k_\ell, k_h] \subset [0, +\infty)$ according to a continuously differentiable cdf G. An equilibrium exists. In equilibrium, the behavior of the opinion leader is characterized by a pair of thresholds (k_0^*, k_1^*) as described in the main text. Furthermore, the expected share of violators is*

$$\overline{v}(\omega \mid e=0) = \overline{v}_b + \frac{\mu}{2-\mu\lambda}\left(\omega - \frac{(1-\mu)\lambda}{2-\lambda}\cdot\frac{G(k_0^*)-G(k_1^*)}{2-G(k_0^*)-G(k_1^*)}\cdot\frac{\psi}{3}\right)$$
$$\overline{v}(\omega \mid e=1) = \overline{v}_b + \frac{\mu}{2-\mu\lambda}\left(\omega + \frac{(1-\mu)\lambda}{2-\lambda}\cdot\frac{G(k_0^*)-G(k_1^*)}{2-G(k_0^*)-G(k_1^*)}\cdot\frac{\psi}{3}\right)$$

The impact of the opinion leader is increasing in $G(k_0^*) - G(k_1^*)$, namely the probability mass with which the ideological strength of the opinion leader is in-between the thresholds k_1^* and k_0^*.

Proof. Assume that in equilibrium there is some information transmission. This happens if a positive mass of opinion leaders play $e = 0$ after signal $s = 0$ and $e = 1$ after signal $s = 1$. Suppose individuals believe the opinion leader behaves according to the threshold strategies (k_0^*, k_1^*) defined in the main text. By Bayes rule, we have:

$$\mathbb{E}[\omega \mid e=0, k_0^*, k_1^*] = -\frac{G(k_0^*)-G(k_1^*))}{2-G(k_0^*)-G(k_1^*)}\cdot\frac{\psi}{3} \tag{C-1}$$
$$\mathbb{E}[\omega \mid e=1, k_0^*, k_1^*] = \frac{G(k_0^*)-G(k_1^*))}{2-G(k_0^*)-G(k_1^*)}\cdot\frac{\psi}{3} \tag{C-2}$$

From the opinion leader's point of view, the expected share of violators is thus equal to

$$\mathbb{E}[\overline{v}(\omega) \mid s] = \begin{cases} \overline{v}_b - \frac{\mu}{2-\mu\lambda}\left(1 + \frac{(1-\mu)\lambda}{2-\lambda}\cdot\frac{G(k_0^*)-G(k_1^*)}{2-G(k_0^*)-G(k_1^*)}\right)\frac{\psi}{3} & \text{if } s=0 \\ \overline{v}_b + \frac{\mu}{2-\mu\lambda}\left(1 + \frac{(1-\mu)\lambda}{2-\lambda}\cdot\frac{G(k_0^*)-G(k_1^*)}{2-G(k_0^*)-G(k_1^*)}\right)\frac{\psi}{3} & \text{if } s=1 \end{cases} \tag{C-3}$$

Opinion leaders with ideological strengths equal to the cutoffs (k_0^*, k_1^*) must be indifferent between endorsing the violation of the norm and not doing so. The thresholds thus jointly solve:

$$k_0^* = 1 - \overline{v}_b + \frac{\mu}{2-\mu\lambda}\left(1 + \frac{(1-\mu)\lambda}{2-\lambda} \cdot \frac{G(k_0^*) - G(k_1^*)}{2 - G(k_0^*) - G(k_1^*)}\right)\frac{\psi}{3} \tag{C-4}$$

$$k_1^* = 1 - \overline{v}_b - \frac{\mu}{2-\mu\lambda}\left(1 + \frac{(1-\mu)\lambda}{2-\lambda} \cdot \frac{G(k_0^*) - G(k_1^*)}{2 - G(k_0^*) - G(k_1^*)}\right)\frac{\psi}{3} \tag{C-5}$$

Obviously, the right hand side of (C-4) is greater than the right hand side of (C-5). Also note that, the right-hand side of (C-4) is bounded above by $k = \frac{1}{2-\lambda} + \frac{\mu}{2-\lambda}\frac{\psi}{3}$. Equations (C-4) and (C-5) define a system of two equations in two unknowns. The right-hand side of this system is a continuous function that maps a convex and compact space, $[0, \max k_h, k] \times [0, \max k_h, k]$ into itself. By the Brouwer fixed point theorem, this system has a solution. Every pair (k_0^*, k_1^*) that satisfies the system is a solution. The previous reasoning immediately implies that $k_0^* \geqslant k_1^*$. Equation (C-3) implies that the impact of the opinion leader's endorsement is larger when the probability mass of opinion leaders with ideological strength in-between k_0^* and k_1^* is larger.

Suppose now that all opinion leaders play $e = 1$ regardless of the signal they receive (i.e., $G(k_0^*) = G(k_1^*) = 0$). In this case, the opinion leader's endorsement has no impact: $\mathbb{E}[\omega \mid e = 1] = 0$. Because the payoff of the opinion leader from endorsing the violation is increasing in k, the endorsement is optimal for all opinion leaders, if and only if it is optimal when (i) the opinion leader has the lowest possible ideological strength, k_ℓ, and (ii) she receives signal $s = 0$. By equation (C-4), this happens when

$$k_\ell \geqslant 1 - \frac{1-\lambda}{2-\lambda} + \frac{\mu\psi}{3(2-\mu\lambda)} = k^\dagger.$$

Next, suppose that all opinion leaders play $e = 0$ regardless of the signal they receive (i.e., $G(k_0^*) = G(k_1^*) = 1$). In this case, we cannot pin down the equilibrium impact of an endorsement. Nonetheless, $e = 0$ is optimal for every impact the endorsement could have, if and only if it is optimal when (i) the impact of an endorsement is maximal, (ii) the opinion leader has the highest possible ideological strength, k_h, and (iii) she receives signal $s = 1$. By equation (C-5), this happens when

$$k_h \leqslant 1 - \frac{1-\lambda}{2-\lambda} - \frac{\mu}{(2-\mu\lambda)}\left(1 + \frac{(1-\mu)\lambda}{2-\lambda}\right)\frac{\psi}{3} = \underline{k}.$$

□

D Informativeness of the Opinion Leader Signal

Proposition D.1. *Let the signal structure be as in equation* (12)*. Then, the share of violators is*

$$\overline{v}(\omega \mid e=0) = \overline{v}_b + \frac{\mu}{2-\mu\lambda}\left[\omega - \frac{(1-\mu)\lambda}{2-\lambda}\cdot\frac{\eta(1)-\eta(0)}{\eta(1)+\eta(0)}\cdot\frac{\psi}{3}\cdot(1-\tau)\right]$$
$$\overline{v}(\omega \mid e=1) = \overline{v}_b + \frac{\mu}{2-\mu\lambda}\left[\omega + \frac{(1-\mu)\lambda}{2-\lambda}\cdot\frac{\eta(1)-\eta(0)}{\eta(1)+\eta(0)}\cdot\frac{\psi}{3}\cdot(1-\tau)\right],$$

where $\overline{v}_b = \frac{1-\lambda}{2-\lambda}$ *is the baseline share of violators when there is no social change (i.e.,* $\mu = 0$*). The impact of the opinion leader's endorsement is decreasing in* τ*. The room for information transmission is also decreasing in* τ*:* $\underline{k}(\tau)$ *increases with* τ*, while* $\overline{k}(\tau)$ *and* $k^{\dagger}(\tau)$ *decrease with* τ*.*

Proof. Let the opinion leader play endorsement strategy $(\eta(0), \eta(1)) \in [0,1]^2$. Individuals' posterior beliefs about ω are:

$$f(\omega \mid e=0) = \frac{1}{2\psi} - (1-\tau)\cdot\frac{\eta(1)-\eta(0)}{2-\eta(1)-\eta(0)}\cdot\frac{\omega}{2\psi^2}$$
$$f(\omega \mid e=1) = \frac{1}{2\psi} + (1-\tau)\cdot\frac{\eta(1)-\eta(0)}{\eta(1)+\eta(0)}\cdot\frac{\omega}{2\psi^2},$$

so that:

$$\mathbb{E}[\omega \mid e=0] = -\frac{\eta(1)-\eta(0)}{2-\eta(1)-\eta(0)}\cdot\frac{\psi}{3}\cdot(1-\tau)$$
$$\mathbb{E}[\omega \mid e=1] = \frac{\eta(1)-\eta(0)}{\eta(1)+\eta(0)}\cdot\frac{\psi}{3}\cdot(1-\tau).$$

If we plug these expectations in the share of norm-violators (see Proposition 1), we obtain the expressions in the statement of the proposition. It is immediate to see that the impact of the opinion leader's endorsement is decreasing in τ.

Now, consider the existence of a fully informative equilibrium, $\eta(0) = 0$ and $\eta(1) = 1$. In such an equilibrium, $\mathbb{E}[\omega \mid s=0] = -\frac{\psi}{3}\cdot(1-\tau)$ and $\mathbb{E}[\omega \mid s=1] = \frac{\psi}{3}\cdot(1-\tau)$.

Credibility constraints thus become:

$$k - \left(\frac{1}{2-\lambda} - \frac{\mu}{2-\mu\lambda} \left(\frac{(1-\mu)\lambda}{2-\lambda} + 1 \right) \cdot \frac{\psi}{3} \cdot (1-\tau) \right) \geqslant 0$$
$$0 \geqslant k - \left(\frac{1}{2-\lambda} - \frac{\mu}{2-\mu\lambda} \left(\frac{(1-\mu)\lambda}{2-\lambda} - 1 \right) \cdot \frac{\psi}{3} \cdot (1-\tau) \right)$$

The first constraint holds if and only if

$$k \geqslant \underline{k}(\tau) = \frac{1}{2-\lambda} \left(1 - \frac{\mu\psi}{3} \cdot (1-\tau) \right),$$

while the second one holds if and only if

$$k \leqslant \overline{k}(\tau) = \frac{1}{2-\lambda} \left(1 + \frac{2(1-\lambda)+\mu\lambda}{2-\mu\lambda} \cdot \frac{\mu\psi}{3} \cdot (1-\tau) \right).$$

A fully informative equilibrium thus exists if and only if $k(\tau) \in \left[\underline{k}(\tau), \overline{k}(\tau) \right]$. The lower bound $\underline{k}(\tau)$ is increasing in τ, while the upper bound $\overline{k}(\tau)$ is decreasing in τ.

Next, consider the existence of a partially informative equilibrium, so that $\eta(0) \in (0,1)$ and $\eta(1) = 1$. Credibility constraints for the opinion leader are now given by:

$$k - \left[\frac{1}{2-\lambda} - \frac{\mu}{2-\mu\lambda} \left(\frac{(1-\mu)\lambda}{2-\lambda} \cdot \frac{1-\eta(0)}{1+\eta(0)} + 1 \right) \cdot \frac{\psi}{3} \cdot (1-\tau) \right] \geqslant 0$$
$$0 \geqslant k - \left[\frac{1}{2-\lambda} - \frac{\mu}{2-\mu\lambda} \left(\frac{(1-\mu)\lambda}{2-\lambda} \cdot \frac{1-\eta(0)}{1+\eta(0)} - 1 \right) \cdot \frac{\psi}{3} \cdot (1-\tau) \right].$$

The second constraint pins down $\eta(0)$. $\eta(0) > 0$ if $k > \overline{k}(\tau)$ and $\eta(0) < 1$ if:

$$k < k^{\dagger}(\tau) = \frac{1}{2-\lambda} \left(1 + \frac{2-\lambda}{2-\mu\lambda} \cdot \frac{\mu\psi}{3} \cdot (1-\tau) \right).$$

A partially informative equilibrium exists if and only if $k \in \left(\overline{k}(\tau), k^{\dagger}(\tau) \right)$ and $k^{\dagger}(\tau)$ is decreasing in τ.[18] □

[18] As it was the case in the proof of Proposition 3, $\eta(0) = 0$ and $\eta(1) \in (0,1)$ leads to a non-generic solution.

E Multiple Opinion Leaders

Assume that two opinion leaders exist. The share of violators is still defined by the expression defined in Proposition 1. However, the information available to individuals now includes the endorsement behavior of both opinion leaders. Since opinion leaders move independently, each opinion leader can play an uninformative strategy, a fully informative strategy, or a partially informative strategy.

Consider opinion leader 1 (the analysis for opinion leader 2 is identical and omitted). Conditional on the signal s that she received, her updated beliefs about the state ω are still as in the baseline model:

$$f(\omega \mid s=0) = \frac{1}{2\psi} - \frac{\omega}{2\psi^2} \qquad \text{and} \qquad f(\omega \mid s=1) = \frac{1}{2\psi} + \frac{\omega}{2\psi^2}.$$

Hence, if she received signal $s = 0$, her expectation about ω is $-\frac{\psi}{3}$. Moreover, the probability she assigns to opinion leader 2 having received signal $s = 1$ is given by:

$$\int_{-\psi}^{\psi} \left(\frac{1}{2\psi} - \frac{\omega}{2\psi^2}\right)\left(\frac{1}{2} + \frac{\omega}{2\psi}\right) d\omega = \frac{1}{3}.$$

Instead, if she received signal $s = 1$, the expected value of ω is $\psi/3$ and the probability she assigns to opinion leader 2 also receiving signal $s = 1$ is:

$$\int_{-\psi}^{\psi} \left(\frac{1}{2\psi} + \frac{\omega}{2\psi^2}\right)\left(\frac{1}{2} + \frac{\omega}{2\psi}\right) d\omega = \frac{2}{3}.$$

Clearly, if opinion leader 2 is playing an uninformative equilibrium strategy, opinion leader 1 is in a situation that is analogous to the one characterized in the main text. Thus, the results in Propositions 2 and 3 still apply.

Suppose that opinion leader 2 is playing a fully informative strategy, $(\eta(0), \eta(1)) = (0, 1)$. If opinion leader 1 also plays a fully informative strategy, individuals can face one of 4 possible pairs of endorsement decisions. The expectations of individuals would react to these pairs as summarized by the following table:

1 \ 2	$b_2 = 0$	$b_2 = 1$
$b_1 = 0$	$-\frac{\psi}{2}$	0
$b_1 = 1$	0	$\frac{\psi}{2}$

Table E1: $\mathbb{E}[\omega \mid \cdot]$ given opinion leaders' endorsement decisions.

Hence, if both opinion leaders play a fully informative strategy, the expected expectation of the individuals from the point of view of the opinion leader is $-\frac{\psi}{2} \cdot \frac{2}{3} = -\frac{\psi}{3}$ after signal $s = 0$ and $\frac{\psi}{2} \cdot \frac{2}{3} = \frac{\psi}{3}$ after signal $s = 1$.

The expected payoff of opinion leader 1 when she she has received signal $s = 1$, she believes opinion leader 2 is playing a fully informative strategy, and she chooses $e = 1$ is equal to:

$$k_1 + \left[1 - v^{FI}(\omega \mid e = 1)\right]$$

where $v^{FI}(\omega \mid e = 1)$ is defined in equation (5). Proceeding as in the proof of Proposition 2 we conclude that truthful information transmission after $s = 1$ is incentive compatible for the opinion leader if and only if $k_1 > \underline{k}$. A similar reasoning also implies that truthful information transmission is incentive compatible for the opinion leader when she receives signal $s = 0$ if and only if $k_1 < \overline{k}$.

The same logic implies that if $k_1 \in (\overline{k}, k^{\dagger})$, opinion leader 1 can play a partially informative equilibrium (see the proof of Proposition 3 for details).

Finally, suppose that opinion leader 2 plays a partially informative strategy in which $\eta(0) \in (0, 1)$ and $\eta(1) = 1$. As in the previous case, the law of iterated expectation implies that the expected expectation of individuals from the point of view of opinion leader 1 is equal to the one in the baseline model. Hence, the expected payoff of opinion leader 1 remains unchanged and the bounds we derived in Propositions 2 and 3 still apply.

The share of violators is obtained by replacing the relevant expectations (e.g., the expressions in Table E1) in the expression for $\overline{v}(\omega)$ in Proposition 1.

References

Acemoglu, D. and M. O. Jackson (2015). History, expectations, and leadership in the evolution of social norms. *The Review of Economic Studies 82*(2), 423–456.

Acemoglu, D. and M. O. Jackson (2017). Social norms and the enforcement of laws. *Journal of the European Economic Association 15*(2), 245–295.

Alesina, A., P. Giuliano, and N. Nunn (2013). On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics 128*(2), 469–530.

Alonso, R. and O. Câmara (2016). Persuading voters. *American Economic Review 106*(11), 3590–3605.

Baron, D. P. (2006). Persistent media bias. *Journal of Public Economics 90*(1), 1–36.

Benabou, R. and J. Tirole (2011). Laws and norms. *NBER working paper No. 17579*.

Bernheim, B. D. (1994). A theory of conformity. *Journal of Political Economy 102*(5), 841–877.

Bilancini, E., L. Boncinelli, and J. Wu (2018). The interplay of cultural intolerance and action-assortativity for the emergence of cooperation and homophily. *European Economic Review 102*, 1–18.

Blackwell, D. (1953). Equivalent comparisons of experiments. *The Annals of Mathematical Statistics 24*(2), 265–272.

Block, J., R. Dutta, and D. Levine (2021). Leaders and social norms: On the emergence of consensus or conflict. *Available at SSRN 4513819*.

Bursztyn, L., G. Egorov, and S. Fiorin (2020). From extreme to mainstream: The erosion of social norms. *American Economic Review 110*(11), 3522–48.

Carlsson, M., G. B. Dahl, and D.-O. Rooth (2016). Do politicians change public attitudes? *IZA Discussion Paper No. 10349*.

Chan, J., S. Gupta, F. Li, and Y. Wang (2019). Pivotal persuasion. *Journal of Economic theory 180*, 178–202.

Che, Y.-K., W. Dessein, and N. Kartik (2013). Pandering to persuade. *American Economic Review 103*(1), 47–79.

Chen, D. L., M. Michaeli, and D. Spiro (2021). Legitimizing policy. *Working paper*.

Chen, H. and W. Suen (2021). Radicalism in mass movements: Asymmetric information and endogenous leadership. *American Political Science Review 115*(1), 286–306.

Chiang, C.-F. and B. Knight (2011). Media bias and influence: Evidence from newspaper endorsements. *The Review of economic studies 78*(3), 795–820.

Cowen, T. and D. Sutter (1998). Why only nixon could go to china. *Public Choice 97*(4), 605–615.

Cukierman, A. and M. Tommasi (1998). When does it take a nixon to go to china? *The American Economic Review 88*(1), 180–197.

Fainmesser, I. P. and A. Galeotti (2021). The market for online influence. *American Economic Journal: Microeconomics 13*(4), 332–372.

Friedrichsen, J., T. König, and T. Lausen (2021). Social status concerns and the political economy of publicly provided private goods. *The Economic Journal 131*(633), 220–246.

Gallice, A. and E. Grillo (2020). Economic and social-class voting in a model of redistribution with social concerns. *Journal of the European Economic Association 18*(6), 3140–3172.

Garthwaite, C. and T. J. Moore (2012, 02). Can Celebrity Endorsements Affect Political Outcomes? Evidence from the 2008 US Democratic Presidential Primary. *The Journal of Law, Economics, and Organization 29*(2), 355–384.

Gentzkow, M. and J. M. Shapiro (2006). Media bias and reputation. *Journal of Political Economy 114*(2), 280–316.

Gerardi, D., E. Grillo, and I. Monzón (2022). The perils of friendly oversight. *Journal of Economic Theory*, 105500.

Gratton, G. (2014). Pandering and electoral competition. *Games and Economic Behavior 84*, 163–179.

Grosjean, P. A., F. Masera, and H. Yousaf (2021). Whistle the racist dogs: Political campaigns and police stops.

Gulotty, R. and Z. Luo (2021). Fire alarm fatigue: How politicians evade accountability. *Available at SSRN 3815391*.

Hermalin, B. E. (1998). Toward an economic theory of leadership: Leading by example. *American Economic Review*, 1188–1206.

Hinnosaar, M. and T. Hinnosaar (2021). Influencer cartels. *Available at SSRN*.

Hopkins, E. and T. Kornienko (2004). Running to keep in the same place: Consumer choice as a game of status. *American Economic Review 94*(4), 1085–1107.

Jackson, M. O. and X. Tan (2013). Deliberation, disclosure of information, and voting. *Journal of Economic Theory 148*(1), 2–30.

Levine, D. K., S. Modica, A. Rustichini, et al. (2022). Cooperating through leaders. *Working paper*.

Levy, G. and R. Razin (2015). Preferences over equality in the presence of costly income sorting. *American Economic Journal: Microeconomics 7*(2), 308–37.

Loeper, A., J. Steiner, and C. Stewart (2014). Influential opinion leaders. *The Economic Journal 124*(581), 1147–1167.

Maskin, E. and J. Tirole (2019). Pandering and pork-barrel politics. *Journal of Public Economics 176*, 79–93.

Michaeli, M. and D. Spiro (2017). From peer pressure to biased norms. *American Economic Journal: Microeconomics 9*(1), 152–216.

Mitchell, M. (2021). Free ad (vice): Internet influencers and disclosure regulation. *The RAND Journal of Economics 52*(1), 3–21.

Morelli, M. and R. Van Weelden (2013). Ideology and information in policymaking. *Journal of Theoretical Politics 25*(3), 412–439.

Morris, S. (2001). Political correctness. *Journal of Political Economy 109*(2), 231–265.

Mullainathan, S. and A. Shleifer (2005). The market for news. *American Economic Review 95*(4), 1031–1053.

Müller, K. and C. Schwarz (2020). From hashtag to hate crime: Twitter and anti-minority sentiment. *Available at SSRN 3149103*.

Ottaviani, M. and P. N. Sørensen (2006a). Professional advice. *Journal of Economic Theory 126*(1), 120–142.

Ottaviani, M. and P. N. Sørensen (2006b). Reputational cheap talk. *The Rand Journal of Economics 37*(1), 155–175.

Prat, A. and D. Strömberg (2013). The political economy of mass media. *Advances in Economics and Econometrics 2*, 135.

Prato, C. and I. R. Turner (2022). Institutional foundations of the power to persuade. *Center for Open Science SocArXiv No. 4w9af*.

Schnakenberg, K. E. (2015). Expert advice to a voting body. *Journal of Economic Theory 160*, 102–113.

Schnakenberg, K. E. (2017). Informational lobbying and legislative voting. *American Journal of Political Science 61*(1), 129–145.