White, Male, and Angry:
A Reputation-based Rationale for Backlash
SUPPLEMENTARY MATERIAL

Stephane Wolton

A Proofs of formalised argument

A.1 Preliminary observations

I detail one of the assumptions mentioned in the text regarding the relationship between the thresholds and the elite size. Under the assumption that all luck shocks are realised for a given ability level, the thresholds for the dominant and dominated groups must satisfy (denoting $\mathbb{E}_\epsilon(\cdot)$ the expectation with respect to the luck shock):

$$e = \alpha \mathbb{E}_\epsilon\big(1 - F(E_D - \widetilde{\epsilon})\big) + (1-\alpha)\mathbb{E}_\epsilon\big(1 - F(E_d - \widetilde{\epsilon})\big) \tag{A.1}$$

Equation A.1 defines the relationship between the two thresholds. This will prove helpful for the proof of Proposition 1.

In the main text, I also state that all individuals remain uncertain about the value of the threshold $\widetilde{E}_D$ after observing their ability and their social status. Formally, this is equivalent to stating that for all $\theta^i \in [\underline{\theta}, \overline{\theta}]$ and for all $s^i \in \{0,1\}$, there does not exist $\widetilde{E}_D \in [\underline{E}_D, \overline{E}_D]$ such that $Pr(\widetilde{E}_D = E_D | s^i, \theta^i) = 1$. Notice that the continuity of all distribution functions and Bayes' rule guarantee that the posterior distributions of the threshold $\widetilde{E}_D$ are continuous over $[\underline{E}_D, \overline{E}_D]$.

To prove some results, I amend the notation in the text and consider E_D as drawn from distribution $\Gamma_D \equiv \Gamma$ and E_d drawn from distribution Γ_d over some $[\underline{E}_d, \overline{E}_d]$ with the two bounds defined by Equation A.1 (for $E_D = \overline{E}_D$ and $E_D = \underline{E}_D$, respectively). Recall that $\widetilde{\cdot}$ denotes random variable and quantity without tilde denotes actual realization. If there is no risk of ambiguity, I also denote $\int_{\widetilde{X}}$ to denote the integral over the whole support of random variable $\widetilde{X}$.

As noted in the text, social reputation is the only interesting quantity here. The average opinion takes value:

$$\begin{aligned}
&\mathbb{E}_{-i}(\widetilde{\theta^i} | g^i, s^i) \\
&= \int_{\widetilde{\theta^j}} \sum_{s \in \{-1,1\},} \sum_{g \in \{D,d\}} \left(\int_{\widetilde{E}} \int_{\widetilde{\epsilon^i}} \mathbb{E}(\widetilde{\theta^i} | g^i, s^i, \widetilde{\epsilon^i}, \widetilde{E}) d\Lambda(\widetilde{\epsilon^i}) d\Gamma_{g^i}(\widetilde{E} | \widetilde{\theta^j}, s^j, g^j = g) \right) \\
&\times P(s^j = s | \widetilde{\theta^j}, g^j = g, E_D) P(g) dF(\widetilde{\theta^j})
\end{aligned}$$

The average opinion consists of the expected ability for a given realization of the luck shock and the threshold given that $s^i = 1 \iff \theta^i + \epsilon^i \geq E_g$. No one observes the luck shock so individuals take into account all possible realizations of the luck shock $\int_{\widetilde{\epsilon^i}} \cdot d\Lambda(\widetilde{\epsilon^i})$. They also take into account the possible realizations of the threshold for individual i from group g^i given what they learned from their own achievements: $\int_{\widetilde{E}} \cdot d\Gamma_{g^i}(\widetilde{E} | \widetilde{\theta^j}, s^j, g^j = g)$. The average opinion is then a function of the proportion of individuals in each social status for each group given the abilities of these individuals: $\int_{\widetilde{\theta^j}} \sum_{s \in \{-1,1\},} \sum_{g \in \{D,d\}} \cdot P(s^j = s | \widetilde{\theta^j}, g^j = g, E_D) P(g) dF(\widetilde{\theta^j})$. Notice that the average opinion depends on the actual proportion of individuals in each group. As such, it depends on the realized threshold for group D and henceforth group d, as noted above (this is a classic case of every individual being wrong, but the average opinion being correct). Using the usual calculations from the law of iterated expectations, I obtain:

$$\mathbb{E}_{-i}(\widetilde{\theta^i}|g^i,s^i) = \int_{\widetilde{E}}\int_{\widetilde{\epsilon^i}} \mathbb{E}(\widetilde{\theta^i}|g^i,s^i,\widetilde{\epsilon^i},\widetilde{E})d\Lambda(\widetilde{\epsilon^i})d\Gamma_{g^i}(\widetilde{E}|E_D)$$

Or, in other words, the average opinion in the population is:

$$\mathbb{E}_{-i}(\widetilde{\theta^i}|g^i,s^i) = \int_{\widetilde{\epsilon^i}} \mathbb{E}(\widetilde{\theta^i}|g^i,s^i,\widetilde{\epsilon^i},E_D)d\Lambda(\widetilde{\epsilon^i})$$

Individual i, however, cannot compute the average opinion in the population since it would require that they know the actual realization of thresholds for the different groups, which I assume they do not. As such, as noted in the text, the relevant quantity is their expectation of how others perceive them, which I have denoted by $\theta^*_g(s^i,\theta^i) \equiv \mathbb{E}^i_{-i}(\widetilde{\theta}|g,s^i,\theta^i)$. To compute this expectation, they form a belief about the realization of the threshold E given their information (status, group, and ability). As such, I obtain:

$$\theta^*_g(s^i,\theta^i) = \int_{\widetilde{E}}\int_{\widetilde{\epsilon}} \mathbb{E}(\widetilde{\theta}|g,s^i,\widetilde{\epsilon},\widetilde{E})d\Lambda(\widetilde{\epsilon})d\Gamma_g(\widetilde{E}|s^i,\theta^i) \tag{A.2}$$

Using this observation and the meritocratic nature of the system, the social reputations take value:

$$\theta^*_g(1,\theta^i) = \int_{\widetilde{E}}\int_{\widetilde{\epsilon}} \mathbb{E}(\widetilde{\theta}|\widetilde{\theta} \geq \widetilde{E}-\widetilde{\epsilon})d\Lambda(\widetilde{\epsilon})d\Gamma_g(\widetilde{E}|1,\theta^i) \tag{A.3}$$

$$\theta^*_g(0,\theta^i) = \int_{\widetilde{E}}\int_{\widetilde{\epsilon}} \mathbb{E}(\widetilde{\theta}|\widetilde{\theta} \leq \widetilde{E}-\widetilde{\epsilon})d\Lambda(\widetilde{\epsilon})d\Gamma_g(\widetilde{E}|0,\theta^i) \tag{A.4}$$

With these expressions, we can prove Lemma 1.

A.2 Proofs of baseline model

Proof of Lemma 1

To prove the first point, fix some $\widetilde{\epsilon}$ and take any θ^i. Using Equation A.3-Equation A.4, we compare:

$$A = \int_{\widetilde{E}} \mathbb{E}(\widetilde{\theta}|\widetilde{\theta} \geq \widetilde{E}-\widetilde{\epsilon})d\Gamma_g(\widetilde{E}|1,\theta^i) \quad \text{and}$$

$$B = \int_{\theta^i+\widetilde{\epsilon^i}}^{\overline{E}_g} \mathbb{E}(\widetilde{\theta}|\widetilde{\theta} \leq \widetilde{E}-\widetilde{\epsilon})d\Gamma_g(\widetilde{E}|0,\theta^i)$$

By properties of conditional expectations, $\mathbb{E}(\widetilde{\theta}|\widetilde{\theta} \geq \widetilde{E}-\widetilde{\epsilon}) > \mathbb{E}(\widetilde{\theta}|\widetilde{\theta} \geq \underline{E}_g-\widetilde{\epsilon})$ for all $\widetilde{E} > \underline{E}_g$ and $\mathbb{E}(\widetilde{\theta}|\widetilde{\theta} \leq \widetilde{E}-\widetilde{\epsilon}) < \mathbb{E}(\widetilde{\theta}|\widetilde{\theta} \leq \overline{E}_g-\widetilde{\epsilon})$ for all $\widetilde{E} < \overline{E}_g$. Hence,

$$A > \int_{\widetilde{E}} \mathbb{E}(\widetilde{\theta}|\widetilde{\theta} \geq \underline{E}_g-\widetilde{\epsilon})d\Gamma_g(\widetilde{E}|1,\theta^i) = \mathbb{E}(\widetilde{\theta}|\widetilde{\theta} \geq \underline{E}_g-\widetilde{\epsilon}) \quad \text{and}$$

$$B < \int_{\widetilde{E}} \mathbb{E}(\widetilde{\theta}|\widetilde{\theta} \leq \overline{E}_g-\widetilde{\epsilon})d\Gamma_g(\widetilde{E}|0,\theta^i) = \mathbb{E}(\widetilde{\theta}|\widetilde{\theta} \leq \overline{E}_g-\widetilde{\epsilon})$$

With the second equality on both lines following from the expectations not depending on $\widetilde{E}$. Further, by properties of conditional expectations, $\mathbb{E}(\widetilde{\theta}|\widetilde{\theta} \leq \overline{E}_g-\widetilde{\epsilon}) \leq \mathbb{E}(\widetilde{\theta}|\underline{E}_g-\widetilde{\epsilon} \leq \widetilde{\theta} \leq \overline{E}_g-\widetilde{\epsilon})$ and $\mathbb{E}(\widetilde{\theta}|\widetilde{\theta} \geq \underline{E}_g-\widetilde{\epsilon}) \geq \mathbb{E}(\widetilde{\theta}|\underline{E}_g-\widetilde{\epsilon} \leq \widetilde{\theta} \leq \overline{E}_g-\widetilde{\epsilon})$. Hence, we obtain $A > B$. Integrating over all possible $\widetilde{\epsilon}$, we obtain that $\theta^*_g(1,\theta^i) > \theta^*_g(0,\theta^i)$ as claimed.

To prove the second point, consider first the function

$$H(E) = \int_{\widetilde{\epsilon}} \frac{\int_{E-\widetilde{\epsilon}}^{\overline{\theta}} \widetilde{\theta} dF(\widetilde{\theta})}{1-F(E-\widetilde{\epsilon})} d\Lambda(\widetilde{\epsilon})$$

Notice that (after re-arranging):

$$H'(E) = \int_{-\bar{\epsilon}}^{\bar{\epsilon}} \frac{f(E-\widetilde{\epsilon})}{1-F(E-\widetilde{\epsilon})} \left(\frac{\int_{E-\widetilde{\epsilon}}^{\overline{\theta}} \widetilde{\theta} dF(\widetilde{\theta})}{1-F(E-\widetilde{\epsilon})} - (E-\widetilde{\epsilon}) \right) d\Lambda(\widetilde{\epsilon}) > 0$$

Then, notice that $\Gamma_g(\widetilde{E}|1,\theta^i) = \int_{\widetilde{\epsilon^i}} \Gamma_g(\widetilde{E}|\theta^i + \widetilde{\epsilon^i} \geq \widetilde{E})$ and $\Gamma_g(\widetilde{E}|0,\theta^i) = \int_{\widetilde{\epsilon^i}} \Gamma_g(\widetilde{E}|\theta^i + \widetilde{\epsilon^i} \leq \widetilde{E})$.

Suppose that individual i belongs to the elite. Fix $\widetilde{\epsilon^i}$ and denote $T^i(\theta^i) = \min\{\widetilde{\epsilon^i} + \theta^i, \overline{E}_g\}$. For all θ^i, $\Gamma_g(\widetilde{E}|\widetilde{E} \leq \theta^i + \widetilde{\epsilon^i}) = \frac{\Gamma_g(\widetilde{E})}{\Gamma_g(T^i(\theta^i))}$ for all $\widetilde{E} \in [\underline{E}_g, T^i(\theta^i)]$, which is decreasing with θ^i (strictly if $T(\theta^i) = \widetilde{\epsilon^i} + \theta^i$), and 1 for $\widetilde{E} \geq T(\theta^i)$. Hence, for all $\theta^i_h > \theta^i_l$, on the union of their support (i.e., $[\underline{E}_g, T(\theta^i_h)]$, we obtain that $\Gamma_g(\widetilde{E}|\widetilde{E} \leq \theta^i_h + \widetilde{\epsilon^i}) \leq \Gamma_g(\widetilde{E}|\widetilde{E} \leq \theta^i_l + \widetilde{\epsilon^i})$ with strict inequality if the union of the support is not empty and $T(\theta^i_l) = \widetilde{\epsilon^i} + \theta^i_l$. Hence, $\Gamma_g(\widetilde{E}|\widetilde{E} \leq \theta^i_h + \widetilde{\epsilon^i})$ "first order stochastic dominates" $\Gamma_g(\widetilde{E}|\widetilde{E} \leq \theta^i_l + \widetilde{\epsilon^i})$, strictly if the union of the support is not an empty interval and $T(\theta^i_l) = \widetilde{\epsilon^i} + \theta^i_l$.[1]
Using the properties of $H(E)$ above, we then obtain for all $\theta^i_h > \theta^i_l$:

$$\int_{\widetilde{E}} H(\widetilde{E}) d\Gamma_g(\widetilde{E}|\widetilde{E} \leq \theta^i_h + \widetilde{\epsilon^i}) \geq \int_{\widetilde{E}} H(\widetilde{E}) d\Gamma_g(\widetilde{E}|\widetilde{E} \leq \theta^i_l + \widetilde{\epsilon^i}),$$

with strict inequality if the union of the support is not empty and $T(\theta^i_l) = \widetilde{\epsilon^i} + \theta^i_l$.

Integrating over all possible $\widetilde{\epsilon^i}$, we obtain:

$$\theta^*_g(1,\theta^i_h) = \int_{\widetilde{\epsilon^i}} \int_{\widetilde{E}} H(\widetilde{E}) d\Gamma_g(\widetilde{E}|\widetilde{E} \leq \theta^i_h + \widetilde{\epsilon^i}) d\Lambda(\widetilde{\epsilon^i}) > \int_{\widetilde{\epsilon^i}} \int_{\widetilde{E}} H(\widetilde{E}) d\Gamma_g(\widetilde{E}|\widetilde{E} \leq \theta^i_l + \widetilde{\epsilon^i}) d\Lambda(\widetilde{\epsilon^i}) = \theta^*_g(1,\theta^i_l)$$

We can apply a similar reasoning for an individual i who belongs to the non-elite group after making two observations. First, $\Gamma_g(\widetilde{E}|\widetilde{E} \geq \theta^i_h + \widetilde{\epsilon^i}) \leq \Gamma_g(\widetilde{E}|\widetilde{E} \geq \theta^i_l + \widetilde{\epsilon^i})$ on the union of their support (with strict inequality when the union is not empty and $\theta^i_h + \widetilde{\epsilon^i} > \underline{E}_g$). Second, the function $J(E) = \int_{\widetilde{\epsilon}} \frac{\int_{E-\widetilde{\epsilon}}^{\underline{\theta}} \widetilde{\theta} dF(\widetilde{\theta})}{1-F(E-\widetilde{\epsilon})} d\Lambda(\widetilde{\epsilon})$ is strictly increasing in E. □

Proof of Proposition 1

I focus first on the dominant group D and discuss the disadvantaged group below. The proof for a member of the dominant group proceeds in five steps.
Step 1: Recall that the conditional pdfs of z satisfy the MLRP, for all $z > z'$. Using Milgrom's (1981) Proposition 1 (p.383), which holds for any non degenerate CDF (our assumption that individuals never perfectly learn the realization of E_D guarantees that the we work with non degenerate distributions), $\Gamma(\cdot|s^i,\theta^i,z)$ first order stochastically dominates $\Gamma(\cdot|s^i,\theta^i,z')$ for all $s^i \in \{0,1\}$ and all $z > z'$, $\theta^i \in [\underline{\theta}, \overline{\theta}]$.
Step 2: We show that Step 1 implies that $\theta^*_D(s^i,\theta^i|z)$ is strictly increasing with z. Denote as above $H(E) = \int_{\widetilde{\epsilon}} \frac{\int_{E-\widetilde{\epsilon}}^{\overline{\theta}} \widetilde{\theta} dF(\widetilde{\theta})}{1-F(E-\widetilde{\epsilon})} d\Lambda(\widetilde{\epsilon})$. Consider an individual from the dominant group with ability θ^i and who belongs to the elite group. By definition of stochastic dominance, given that $H(E)$ is a strictly increasing function, for all $z > z'$:

$$\theta^*_D(s^i,\theta^i|z) = \int_{\widetilde{E}} H(\widetilde{E}) d\Gamma(\widetilde{E}|s^i,\theta^i,z) > \int_{\underline{E}_D}^{\overline{E}_D} H(\widetilde{E}) d\Gamma(\widetilde{E}|s^i,\theta^i,z') = \theta^*_D(s^i,\theta^i|z')$$

Hence, $\theta^*_D(s^i,\theta^i|z)$ is strictly increasing with z.
Step 3: $\Gamma(\widetilde{E}|s^i,\theta^i)$ (the interim distribution, prior to the public signal z) first order stochastically dominates $\Gamma(\cdot|s^i,\theta^i,\underline{z})$. To see this, suppose it does not. First, suppose that there exists $E \in [\underline{E}_D, \overline{E}_D]$ such that $\Gamma(E|s^i,\theta^i,\underline{z}) < \Gamma(E|s^i,\theta^i)$. Now, since $\Gamma(\widetilde{E}|s^i,\theta^i,\underline{z})$ is first order stochastically dominated by $\Gamma(\widetilde{E}|s^i,\theta^i,z)$ for all $z > \underline{z}$, we must have $\Gamma(E|s^i,\theta^i,z) < \Gamma(E|s^i,\theta^i)$. Then, $\int_{\underline{z}}^{\overline{z}} \Gamma(E|s^i,\theta^i,\widetilde{z}) dZ(\widetilde{z}) > \int_{\underline{z}}^{\overline{z}} \Gamma(E|s^i,\theta^i) dZ(\widetilde{z})$. By

[1] I include quotation marks as first order stochastic dominance supposes that the distributions have the same support. However, these specific truncations are clearly related to it.

the law of total probabilities, $\int_{\underline{z}}^{\overline{z}} \Gamma(E|s^i,\theta^i,\widetilde{z})dZ(\widetilde{z}) = \Gamma(E|s^i,\theta^i)$. Since $\int_{\underline{z}}^{\overline{z}} \Gamma(E|s^i,\theta^i)dZ(\widetilde{z}) = \Gamma(E|s^i,\theta^i)$, we obtain $\Gamma(E|s^i,\theta^i) > \Gamma(E|s^i,\theta^i)$ a contradiction. Now, suppose that for all E, $\Gamma(E|s^i,\theta^i,\underline{z}) = \Gamma(E|s^i,\theta^i)$. Since $\Gamma(\widetilde{E}|s^i,\theta^i,\underline{z})$ is first order stochastically dominated by $\Gamma(\widetilde{E}|s^i,\theta^i,z)$ for all $z > \underline{z}$, there exists E' such as by the same reasoning as above, we obtain $\Gamma(E'|s^i,\theta^i) > \Gamma(E'|s^i,\theta^i)$, a contradiction.
Step 4: by the same reasoning, we can show that $\Gamma(\widetilde{E}|s^i,\theta^i)$ is first order stochastically dominated by $\Gamma(\widetilde{E}|s^i,\theta^i,\overline{z})$. Using this result and $\Gamma(\widetilde{E}|s^i,\theta^i)$ FOSD $\Gamma(\widetilde{E}|s^i,\theta^i,\underline{z})$, we obtain that $\theta^*_D(s^i,\theta^i|\underline{z}) < \theta^*_D(s^i,\theta^i) < \theta^*_D(s^i,\theta^i|\overline{z})$ (again using step 2).
Step 5: Combining the results from Step 2 ($\theta^*_D(s^i,\theta^i|z)$ strictly increasing in z) and from Step 4 ($\theta^*_D(s^i,\theta^i|\underline{z}) < \theta^*_D(s^i,\theta^i) < \theta^*_D(s^i,\theta^i|\overline{z})$) and the theorem of intermediate values, we obtain that there exists a unique $z^0(s^i,\theta^i,D)$ such that $\theta^*_D(s^i,\theta^i) \leq (>) \theta^*_D(s^i,\theta^i|z)$ for all $z \geq (<) z^0(s^i,\theta^i,D)$.
Turning to the disadvantaged group, define the distribution of thresholds for the disadvantaged group as Equation A.1 as:

$$\Gamma_d(E) = 1 - \Gamma\left(v^{-1}\left(\frac{1-e-(1-\alpha)v(E)}{\alpha}\right)\right), \tag{A.5}$$

with $v(E) = \mathbb{E}_\epsilon(F(E-\widetilde{\epsilon}))$ a strictly increasing function in E.
It follows that if $\Gamma(\widetilde{E}|s^i,\theta^i,z)$ first order stochastically dominates $\Gamma(\widetilde{E}|s^i,\theta^i,z')$, then the associated $\Gamma_d(\widetilde{E}|s^i,\theta^i,z)$ is first order stochastically dominated by $\Gamma_d(\widetilde{E}|s^i,\theta^i,z')$. Hence, all the results for the dominant group above are inverted for the disadvantaged group appropriately adapting the notation.
To prove the last point of the proposition, notice that if there exists an uninformative messagez^u, then necessarily $\theta^*_g(s^i,\theta^i|z^u) = \theta^*_g(s^i,\theta^i)$. Since $z^0(s^i,\theta^i,g)$ is unique, it must be that $z^0(s^i,\theta^i,g) = z^u$ for all s^i, θ^i, and g. □

A.3 Proofs of amended model

Proof of Proposition 2

Consider an individual from the dominant group D with ability θ^i. Slightly amending notation, denote $\theta^*_D(s,\Delta)$ the reputation of group-D members with social status $s \in \{0,1\}$ after the threshold has been increased by Δ (notice that as per the above, the ability θ^i only matters to update about the threshold for entry into the elite, since the threshold is now assumed to be known, I omit ability from the notation of social reputation). The expected payoff of this individual is:

$$\begin{aligned} W_D(\theta^i,\Delta) = & \big(1-\Lambda(E_D+\Delta-\theta^i)\big)\big(1+\theta^*_D(1,\Delta)\big) \\ & + \Lambda(E_D+\Delta-\theta^i)\big(0+\theta^*_D(0,\Delta)\big) \end{aligned} \tag{A.6}$$

The first term after the equal sign $(\big(1-\Lambda(E_D+\Delta-\theta^i)\big))$ corresponds to the probability of joining the elite for an individual with ability θ^i: the luck shock must be high enough for individual i to pass the threshold E_D. The second term $(1+\theta^*_D(1,\Delta))$ corresponds to the payoff when in the elite. On the second line, the terms consists of the probability of missing the bar and the payoff when not in the elite.
Assume first that $\Lambda(E_D-\overline{\theta}) > 0$ and $\Lambda(E_D-\underline{\theta}) < 1$ (i.e., even the highest ability individual may fail to join the elite due to bad luck and the lowest ability individual may join the elite thanks to good luck). Taking the derivative with respect to Δ, I obtain:

$$\begin{aligned} \frac{\partial W_D(\theta^i,\Delta)}{\partial \Delta} = & -\lambda(E_D+\Delta-\theta^i)\big(1+\theta^*_D(1,\Delta)-\theta^*_D(0,\Delta)\big) \\ & + \big(1-\Lambda(E_D+\Delta-\theta^i)\big)\frac{\partial \theta^*_D(1,\Delta)}{\partial \Delta} \\ & + \Lambda(E_D+\Delta-\theta^i)\frac{\partial \theta^*_D(0,\Delta)}{\partial \Delta} \end{aligned}$$

Notice that using the proof of Lemma 1,

$$\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta} = \int_{-\bar{\epsilon}}^{\bar{\epsilon}} \frac{f(E_D+\Delta-\widetilde{\epsilon})}{1-F(E_D+\Delta-\widetilde{\epsilon})}\left(\frac{\int_{E_D+\Delta-\widetilde{\epsilon}}^{\overline{\theta}}\widetilde{\theta}dF(\widetilde{\theta})}{1-F(E_D+\Delta-\widetilde{\epsilon})}-(E_D+\Delta-\widetilde{\epsilon})\right)d\Lambda(\widetilde{\epsilon})>0$$

$$\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta} = \int_{-\bar{\epsilon}}^{\bar{\epsilon}} \frac{f(E_D+\Delta-\widetilde{\epsilon})}{F(E_D+\Delta-\widetilde{\epsilon})}\left((E_D+\Delta-\widetilde{\epsilon})-\frac{\int_{\underline{\theta}}^{E_D+\Delta-\widetilde{\epsilon}}\widetilde{\theta}dF(\widetilde{\theta})}{F(E_D+\Delta-\widetilde{\epsilon})}\right)d\Lambda(\widetilde{\epsilon})>0$$

Now consider how the derivative of $W_D(\theta^i,\Delta)$ wrt to Δ varies with ability θ^i:

$$\begin{aligned}\frac{\partial^2 W_D(\theta^i,\Delta)}{\partial\Delta\partial\theta^i} =& \lambda'(E_D+\Delta-\theta^i)\big(1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)\big)\\ &+\lambda(E_D+\Delta-\theta^i)\left(\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}\right)\end{aligned}$$

Rearranging, the sign of $\frac{\partial^2 W_D(\theta^i,\Delta)}{\partial\Delta\partial\theta^i}$ is the same as the sign of

$$\frac{\lambda'(E_D+\Delta-\theta^i)}{\lambda(E_D+\Delta-\theta^i)}+\frac{\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}}{1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)}$$

Since $\frac{\lambda'(\epsilon)}{\lambda(\epsilon)}$ is decreasing with ϵ by assumption, $\frac{\lambda'(\widetilde{E}+\Delta-\theta^i)}{\lambda(\widetilde{E}+\Delta-\theta^i)}$ evaluated at $\Delta=0$ is increasing with θ^i. As a result, there are three cases to consider:

(1) $\frac{\partial^2 W_D(\theta^i,0)}{\partial\Delta\partial\theta^i}$ is negative for all θ^i;

(2) $\frac{\partial^2 W_D(\theta^i,0)}{\partial\Delta\partial\theta^i}$ is positive for all θ^i;

(3) There exists θ^+ such that $\frac{\partial^2 W_D(\theta^i,0)}{\partial\Delta\partial\theta^i}$ is strictly negative for all $\theta^i<\theta^+$ and positive for all $\theta^i>\theta^+$ (zero at $\theta^i=\theta^+$).

In all cases, we can have $\frac{\partial W_D(\theta^i,0)}{\partial\Delta}<0$ for all θ^i, in which cases pick $\theta_D^l<\underline{\theta}$ and $\overline{\theta}<\theta_D^h$, or $\frac{\partial W_D(\theta^i,0)}{\partial\Delta}>0$ for all θ^i, in which case pick $\overline{\theta}<\theta_D^l<\theta_D^h$. On top of this,

- In cases (1) and (3), if there exists a unique solution in $\theta^s\in[\underline{\theta},\overline{\theta}]$ to $\frac{\partial W_D(\theta^i,\Delta)}{\partial\Delta}=0$ such that $\frac{\partial W_D(\theta^i,\Delta)}{\partial\Delta}<0$ for all $\theta^i>\theta^s$, denote $\theta^s=\theta_D^l$ and pick $\theta_D^h>\overline{\theta}$.
- In cases (2) and (3), if there exists a unique solution in $\theta^s\in[\underline{\theta},\overline{\theta}]$ to $\frac{\partial W_D(\theta^i,0)}{\partial\Delta}=0$ such that $\frac{\partial W_D(\theta^i,0)}{\partial\Delta}>0$ for all $\theta^i>\theta^s$, denote $\theta^s=\theta_D^h$ and pick $\theta_D^l<\underline{\theta}$.
- In case (3), if there exists two solution in $\theta_1^s,\theta_2^s\in[\underline{\theta},\overline{\theta}]^2$ to $\frac{\partial W_D(\theta^i,0)}{\partial\Delta}=0$ denote $\theta_1^s=\theta_D^l$ and $\theta_2^s=\theta_D^h$.

This represents all possible cases. In all these cases, we have been able to define θ_D^l and θ_D^h satisfying the conditions of the proposition for the dominant group.

Now, relax the assumption that $\Lambda(E_D-\overline{\theta})>0$ and $\Lambda(E_D-\underline{\theta})<1$. Suppose for example that there exists a unique $\theta^T\in(\underline{\theta},\overline{\theta})$ such that $\Lambda(E_D-\theta^i)=0$ for all $\theta^i\geq\theta^T$ (whereas $\Lambda(E_D-\underline{\theta})<1$). Then, for all $\theta^i>\theta^T$, I obtain:

$$\frac{\partial W_D(\theta^i,0)}{\partial\Delta}=\frac{\partial\theta_D^*(1,0)}{\partial\Delta}>0$$

For all other $\theta^i\leq\theta^T$, a similar reasoning as above applies. Hence, we know have the following possible cases:

- $\frac{\partial W_D(\theta^i,0)}{\partial\Delta}>0$ for all θ^i, in which case pick $\overline{\theta}<\theta_D^l<\theta_D^h$
- There exists a unique solution in $\theta^s\in[\underline{\theta},\overline{\theta}]$ to $\frac{\partial W_D(\theta^i,0)}{\partial\Delta}=0$ such that $\frac{\partial W_D(\theta^i,0)}{\partial\Delta}>0$ for all $\theta^i>\theta^s$, denote $\theta^s=\theta_D^h$ and pick $\theta_D^l<\underline{\theta}$.
- There exists two solution in $\theta_1^s,\theta_2^s\in[\underline{\theta},\overline{\theta}]^2$ to $\frac{\partial W_D(\theta^i,0)}{\partial\Delta}=0$ denote $\theta_1^s=\theta_D^l$ and $\theta_2^s=\theta_D^h$.

We again have been able to define θ^l_D and θ^h_D satisfying the conditions of the proposition for the dominant group.
A similar reasoning applies to the case when $\Lambda(E_D - \underline{\theta}) = 1$.

We then can apply a similar reasoning for the dominated group noting that δ has the opposite effect than Δ for individuals from group d. $\square$

B Robustness of formal results

B.1 Robustness of baseline model results

In the baseline model, I make several assumptions: (1) Individuals do not know the thresholds to join the elite E_d and E_D, (2) Individuals know their ability, (3) Individuals know the distributions of ability, (4) Individuals know the size of the elite, (5) Individuals do not know the composition of the elite. In this appendix, I show that the insights from Proposition 1 are robust to relaxing or changing some of these assumptions. I proceed in several steps. I first show that Proposition 1 holds when assumptions 2 and 4 are relaxed (keeping the other assumption). I then explain how we can still obtain a similar result as in Proposition 1 when individuals know the value of the thresholds, but do not know the distributions of ability. I also highlight how information can negatively affect the dominant group when the composition of the elite is known, but the thresholds and the distributions of ability are not.

These various extensions are meant to illustrate that the key assumption for Proposition 1 to hold is that individuals face some uncertainty about what success/failure means for the way the system works or the composition of society. With a mass of individuals, this requires at least two sources of uncertainty. Indeed, suppose that Assumptions 1-4 hold, but I relax assumption 5. Then, given a fixed elite size, individuals can recover the value of the thresholds. There would not be any uncertainty left and, therefore, no role for information.

Uncertainty about the threshold values

In this subsection, I show that Proposition 1 does not depend on assumptions 2 and 4 above.

Suppose that individuals do not perfectly observe their ability (i.e., relaxing assumption 2). This would only affect how individuals compute their expected reputation. To see that, suppose that citizens receive a signal η^i distributed according to CDF $P(\eta^i|\theta^i)$ and pdf $p(\eta^i|\theta^i)$. The signal could be fully informative (in which case, $P(\eta^i|\theta^i)$ is a degenerate distribution), completely uninformative (in which case, $p(\eta^i|\theta^i) = p(\eta^i|\theta^{i\prime})$ for all $\eta^i, \theta^i, \theta^{i\prime}$ in their relevant supports), or anything in between. The expected reputation then becomes using Equation A.2:

$$\mathbb{E}^i_{-i}(\widetilde{\theta}|g^i, s^i, \eta^i) = \int_{\widetilde{E}} \int_{\widetilde{\epsilon}} \mathbb{E}(\widetilde{\theta}|g^i, s^i, \widetilde{\epsilon}, \widetilde{E}) d\Lambda(\widetilde{\epsilon}) d\Gamma_g(\widetilde{E}|s^i, g^i, \eta^i)$$

The proof of Proposition 1 then would go through unchanged after appropriately replacing θ^i by η^i. The proof of the first point of Lemma 1 would remain the same as above again. The proof of the second point of Lemma 1 with respect to η^i would hold if we impose the MLRP on the signals.

Suppose instead that the size of the elite e is unknown (relaxing assumption 4). Then, the common prior is that $\widetilde{e}$ is distributed according to CDF $\mathcal{E}$ and strictly positive pdf ε over $[\underline{e}, \overline{e}]$. Since any individual is atomistic, their own success or failure cannot influence their belief about the size of the elite. Hence, using Equation A.2, the social reputation becomes:

$$\mathbb{E}^i_{-i}(\widetilde{\theta}|g^i, s^i, \theta^i) = \int_{\underline{e}}^{\overline{e}} \int_{\underline{E}(\widetilde{e})}^{\overline{E}(\widetilde{e})} \int_{\widetilde{\epsilon}} \mathbb{E}(\widetilde{\theta}|g^i, s^i, \widetilde{\epsilon}, \widetilde{E}) d\Lambda(\widetilde{\epsilon}) d\Gamma_g(\widetilde{E}|s^i, g^i, \eta^i, \widetilde{e}) d\mathcal{E}(\widetilde{e})$$

Notice that if the size of the elite does not affect the possible bound of the thresholds for entering the elite or the distribution of the thresholds, then uncertainty about the elite size does not matter. I suppose that either the bounds or the distribution is affected by the size of the elite. The next step is to note that Proposition 1 is obtained for one particular realisation of e for the case of uncertain elite size. Slightly abusing notation, we can rewrite the expected reputation in the case of a fixed e (Equation A.3 and Equation A.4) as conditional on a particular realization of the elite size:

$$\theta_g^*(1,\theta^i|e) = \int_{\widetilde{\epsilon}} \frac{\int_{\widetilde{E}(e)-\widetilde{\epsilon}}^{\overline{\theta}} \widetilde{\theta} dF(\widetilde{\theta})}{1-F(\widetilde{E}-\widetilde{\epsilon})} d\Lambda(\widetilde{\epsilon}) d\Gamma_g(\widetilde{E}|1,\theta^i,e)$$

$$\theta_g^*(0,\theta^i|e) = \frac{\int_{\underline{\theta}}^{\widetilde{E}(e)-\widetilde{\epsilon}} \widetilde{\theta} dF(\widetilde{\theta})}{F(\widetilde{E}-\widetilde{\epsilon})} d\Lambda(\widetilde{\epsilon}) d\Gamma_g(\widetilde{E}|0,\theta^i,e)$$

The expected reputation with uncertain elite size is then:

$$\theta_g^*(s^i,\theta^i) = \int_{\widetilde{e}} \theta_g^*(s^i,\theta^i|\widetilde{e}) d\mathcal{E}(\widetilde{e}), \text{for all } s^i \in \{0,1\}$$

Since $\varepsilon(\widetilde{e}) > 0$, the integration over $\widetilde{e}$ preserves inequalities and Lemma 1 and Proposition 1 hold when the size of the elite is uncertain.

Uncertainty about the distributions of ability

In this subsection, I take an alternative approach to the baseline model. I assume that individuals know E_D and E_d (modifying assumption 1 above). I suppose that they are uncertain about the distribution of abilities in both groups (modifying assumption 3 above). I keep all the other assumptions as in the baseline model (i.e., individuals know their ability, the size of the elite, but do not know the composition of the elite).
Denote $\mathcal{F}_g$ the set of possible pdf f_g of ability over $[\underline{\theta},\overline{\theta}]$ for group g and $\widetilde{f}_g$ the random variable over the possible realization of f_g. Due to the difficulties of working with second-order uncertainty (uncertainty about the distributions of random variable), I make a few assumptions for tractability. First, I assume that the set $\mathcal{F}_g$ contains countably many elements: $\mathcal{F}_g = \{f_g^1, f_g^2, f_g^3, ...\}$. I denote the cardinality of $\mathcal{F}_g$ by n (note that we can approximate the continuous case by taking n to infinity) and assume that the last element in $\mathcal{F}_g$ is f_g^n. Second, I assume that distributions are ranked in the sense of strict monotone likelihood ratio property. That is, I order the distribution so that $f_g^k > f_g^j \iff$ for all $\theta_h^i, \theta_l^i \in [\underline{\theta},\overline{\theta}]^2$ with $\theta_h^i > \theta_l^i$, $\frac{f_g^k(\theta_h^i)}{f_g^k(\theta_l^i)} > \frac{f_g^j(\theta_h^i)}{f_g^j(\theta_l^i)}$ (I also sometimes state results only focusing on the superscripts of the pdfs since it is equivalent).[2] The prior distribution satisfies: $Pr(\widetilde{f}_g = f_g^j) = \pi_g^j$ for $g \in \{d,D\}$, with $\pi_g^j > 0$ for all $j \in \{1,...,n\}$. Finally, all distributions in $\mathcal{F}_g$ satisfy the conditions in the main text (i.e., all pdfs are continuous).
I assume that for each realized distribution in $\mathcal{F}_D$ there is an appropriate realized distribution in $\mathcal{F}_d$ so that the following equation holds:

$$e = \alpha \mathbb{E}_\epsilon\big(1 - F_D^h(E_D - \widetilde{\epsilon})\big) + (1-\alpha)\mathbb{E}_\epsilon\big(1 - F_d^k(E_d - \widetilde{\epsilon})\big) \tag{B.1}$$

As the MLRP implies first order stochastic dominance, Equation B.1 directly implies that a higher realized distribution for the dominant group (i.e., a higher superscript) means a lower realized distribution for the disadvantaged group (i.e., a lower superscript).
The social reputation is again the only quantity of interest and I denote it by: $\theta_g^\dagger(s^i,\theta^i)$ for an individual from group g with status s^i and ability θ^i. Denote $\mu_g^k(\theta^i) = Pr(\widetilde{f}_g = f^k|\theta^i)$ the posterior that the probability density distribution of ability is f^k after individual i observes their ability θ^i. In this case, using the same steps as in Online Appendix A.1, the social reputation is:

$$\theta_g^\dagger(s^i,\theta^i) = \sum_{k=1}^{n} \int_{\widetilde{\epsilon}} \mathbb{E}(\widetilde{\theta}|E_D, s^i, \widetilde{\epsilon}, f_g^k) d\Lambda(\widetilde{\epsilon}) \mu_g^k(\theta^i) \tag{B.2}$$

Rather than integrating over possible realization of the thresholds given θ^i as in Equation A.2, I now sum over possible realization of the distributions of ability given θ^i. With this, we can state the equivalent to Lemma 1.

[2] Ranking in term of first order stochastic dominance would be enough to prove an equivalent result to Proposition 1. The stronger assumption I state is sufficient to recover a similar result as Lemma 1.

Lemma B.1. *Elite members have higher expected reputation than non-elite members: for all $\theta^i \in [\underline{\theta}, \overline{\theta}]$, $\theta_g^\dagger(1, \theta^i) > \theta_g^\dagger(0, \theta^i)$ for all $g \in \{d, D\}$.*
An individual's social reputation increases with their own ability: for all $\theta_h^i, \theta_l^i \in [\underline{\theta}, \overline{\theta}]^2$ satisfying $\theta_h^i > \theta_l^i$ $\theta_g^\dagger(s^i, \theta_h^i) > \theta_g^\dagger(s^i, \theta_l^i)$ for all $g \in \{d, D\}$ and $s^i \in \{0, 1\}$.

Proof. Notice that Equation B.2 yields $\theta_g^*(1, \theta^i) > \theta_g^*(0, \theta^i)$ since the social status only enters the conditional expectation (the entry of one individual into the elite is independent of the distribution of abilities).
The second point of the lemma requires more work. First, notice that $f(\theta^i | \widetilde{f}_g = f_g^h) = f_g^h(\theta^i)$. Hence, we have $\frac{f(\theta_h^i | \widetilde{f}_g = f_g^h)}{f(\theta_l^i | \widetilde{f}_g = f_g^h)} > \frac{f(\theta_h^i | \widetilde{f}_g = f_g^l)}{f(\theta_l^i | \widetilde{f}_g = f_g^l)}$ for all $\theta_h^i > \theta_l^i$ and all $f_g^h > f_g^l$ (in the order I have defined above).
Second, for all $\theta_h^i > \theta_l^i$, there exists $1 \leq k^0(\theta_h^i, \theta_l^i) < n$ such that $\mu_g^j(\theta_h^i) < (\leq) \mu_g^j(\theta_l^i)$ if $j < (\leq) k^0(\theta_h^i, \theta_l^i)$ and $\mu_g^j(\theta_h^i) > \mu_g^j(\theta_l^i)$ if $j > k^0(\theta_h^i, \theta_l^i)$. To see this, note that $\mu_g^j(\theta^i) = \frac{\pi_g^j f_g^j(\theta^i)}{\sum_{k=1}^n \pi_g^k f_g^k(\theta^i)}$, or equivalently: $\mu_g^j(\theta^i) = \frac{\pi_g^j}{\sum_{k=1}^n \pi_g^k \frac{f_g^k(\theta^i)}{f_g^j(\theta^i)}}$. Hence, $\mu_g^j(\theta_h^i) > \mu_g^j(\theta_l^i)$ if and only if $\sum_{k=1}^n \pi_g^k \frac{f_g^k(\theta_h^i)}{f_g^j(\theta_h^i)} < \sum_{k=1}^n \pi_g^k \frac{f_g^k(\theta_l^i)}{f_g^j(\theta_l^i)}$. Given the MLRP of the pdfs, we necessarily have $\mu_g^1(\theta_h^i) < \mu_g^1(\theta_l^i)$ and $\mu_g^n(\theta_h^i) > \mu_g^n(\theta_l^i)$. Further, if for $h \in \{2, ..., n-1\}$, $\mu_g^h(\theta_h^i) \leq \mu_g^h(\theta_l^i)$ then $\mu_g^j(\theta_h^i) < \mu_g^j(\theta_l^i)$ for all $j < h$. To see that, recall that $\mu_g^h(\theta_h^i) \leq \mu_g^h(\theta_l^i)$ is equivalent to $\sum_{k=1}^n \pi_g^k \frac{f_g^k(\theta_h^i)}{f_g^h(\theta_h^i)} - \sum_{k=1}^n \pi_g^k \frac{f_g^k(\theta_l^i)}{f_g^h(\theta_l^i)} \geq 0$. Now take

$$\begin{aligned}
\sum_{k=1}^n \pi_g^k \frac{f_g^k(\theta_h^i)}{f_g^j(\theta_h^i)} - \sum_{k=1}^n \pi_g^k \frac{f_g^k(\theta_l^i)}{f_g^j(\theta_l^i)} &= \sum_{k=1}^n \pi_g^k \frac{f_g^k(\theta_h^i)}{f_g^h(\theta_h^i)} \frac{f_g^h(\theta_h^i)}{f_g^j(\theta_h^i)} - \sum_{k=1}^n \pi_g^k \frac{f_g^k(\theta_l^i)}{f_g^h(\theta_l^i)} \frac{f_g^h(\theta_l^i)}{f_g^j(\theta_l^i)} \\
&= \frac{f_g^h(\theta_h^i)}{f_g^j(\theta_h^i)} \sum_{k=1}^n \pi_g^k \frac{f_g^k(\theta_h^i)}{f_g^h(\theta_h^i)} - \frac{f_g^h(\theta_l^i)}{f_g^j(\theta_l^i)} \sum_{k=1}^n \pi_g^k \frac{f_g^k(\theta_l^i)}{f_g^h(\theta_l^i)}
\end{aligned}$$

Since $j < h$ and $\theta_h^i > \theta_l^i$, $\frac{f_g^h(\theta_h^i)}{f_g^j(\theta_h^i)} > \frac{f_g^h(\theta_l^i)}{f_g^j(\theta_l^i)}$ given the ordering of distributions. Hence, I obtain:

$$\sum_{k=1}^n \pi_g^k \frac{f_g^k(\theta_h^i)}{f_g^j(\theta_h^i)} - \sum_{k=1}^n \pi_g^k \frac{f_g^k(\theta_l^i)}{f_g^j(\theta_l^i)} > \frac{f_g^h(\theta_h^i)}{f_g^j(\theta_h^i)} \left(\underbrace{\sum_{k=1}^n \pi_g^k \frac{f_g^k(\theta_h^i)}{f_g^h(\theta_h^i)} - \sum_{k=1}^n \pi_g^k \frac{f_g^k(\theta_l^i)}{f_g^h(\theta_l^i)}}_{\geq 0} \right) > 0$$

A similar reasoning yields that if for $h \in \{2, ..., n-1\}$, $\mu_g^h(\theta_h^i) > \mu_g^h(\theta_l^i)$ then $\mu_g^j(\theta_h^i) > \mu_g^j(\theta_l^i)$ for all $j > h$. Taking all findings together, this implies that there exists a unique $k^0(\theta_h^i, \theta_l^i)$ satisfying $1 \leq k^0(\theta_h^i, \theta_l^i) < n$ such that $\mu_g^j(\theta_h^i) < (\leq) \mu_g^j(\theta_l^i)$ if $j < (\leq) k^0(\theta_h^i, \theta_l^i)$ and $\mu_g^j(\theta_h^i) > \mu_g^j(\theta_l^i)$ if $j > k^0(\theta_h^i, \theta_l^i)$.
With this, we can show that $\theta_g^\dagger(s^i, \theta_h^i) > \theta_g^\dagger(s^i, \theta_l^i)$. Write

$$\begin{aligned}
\theta_g^\dagger(s^i, \theta_h^i) - \theta_g^\dagger(s^i, \theta_l^i) =& \sum_{k=1}^n \int_{\widetilde{\epsilon}} \mathbb{E}(\widetilde{\theta} | E_D, s^i, \widetilde{\epsilon}, f_g^k) d\Lambda(\widetilde{\epsilon}) (\mu_g^k(\theta_h^i) - \mu_g^k(\theta_l^i)) \\
=& \sum_{k=1}^{k^0(\theta_h^i, \theta_l^i)} \int_{\widetilde{\epsilon}} \mathbb{E}(\widetilde{\theta} | E_D, s^i, \widetilde{\epsilon}, f_g^k) d\Lambda(\widetilde{\epsilon}) \overbrace{(\mu_g^k(\theta_h^i) - \mu_g^k(\theta_l^i))}^{\leq 0} \\
&+ \sum_{k=k^0(\theta_h^i, \theta_l^i)+1}^{n} \int_{\widetilde{\epsilon}} \mathbb{E}(\widetilde{\theta} | E_D, s^i, \widetilde{\epsilon}, f_g^k) d\Lambda(\widetilde{\epsilon}) \underbrace{(\mu_g^k(\theta_h^i) - \mu_g^k(\theta_l^i))}_{>0}
\end{aligned}$$

Given the way I order the pdf (according to the MLRP), F^h first order stochastically dominate F^l for all $h > l$ and therefore $\int_{\widetilde{\epsilon}} \mathbb{E}(\widetilde{\theta}|E_D, s^i, \widetilde{\epsilon}, f_g^h) d\Lambda(\widetilde{\epsilon}) > \int_{\widetilde{\epsilon}} \mathbb{E}(\widetilde{\theta}|E_D, s^i, \widetilde{\epsilon}, f_g^l) d\Lambda(\widetilde{\epsilon})$. Hence,

$$\begin{aligned}\theta_g^\dagger(s^i, \theta_h^i) - \theta_g^\dagger(s^i, \theta_l^i) > & \sum_{k=1}^{k^0(\theta_h^i, \theta_l^i)} \int_{\widetilde{\epsilon}} \mathbb{E}(\widetilde{\theta}|E_D, s^i, \widetilde{\epsilon}, f_g^{k^0(\theta_h^i, \theta_l^i)+1}) d\Lambda(\widetilde{\epsilon})(\mu_g^k(\theta_h^i) - \mu_g^k(\theta_l^i)) \\ & + \sum_{k=k^0(\theta_h^i, \theta_l^i)+1}^{n} \int_{\widetilde{\epsilon}} \mathbb{E}(\widetilde{\theta}|E_D, s^i, \widetilde{\epsilon}, f_g^{k^0(\theta_h^i, \theta_l^i)+1}) d\Lambda(\widetilde{\epsilon})(\mu_g^k(\theta_h^i) - \mu_g^k(\theta_l^i)) \\ & = \int_{\widetilde{\epsilon}} \mathbb{E}(\widetilde{\theta}|E_D, s^i, \widetilde{\epsilon}, f_g^{k^0(\theta_h^i, \theta_l^i)+1}) d\Lambda(\widetilde{\epsilon}) \times \sum_{k=1}^{n} (\mu_g^k(\theta_h^i) - \mu_g^k(\theta_l^i)) = 0\end{aligned}$$

□

To think about the effect of public information in this context, I consider a public signal $y \in [\underline{y}, \overline{y}]$ distributed conditional on a distribution of ability in the dominant group f_D according to the pdf and CDF $\psi(y|f_D)$ and $\Psi(y|f_D)$. I suppose that the conditional distributions satisfy the following property for all $y^t, y^b \in [\underline{y}, \overline{y}]^2$ with $y^t > y^b$ and for all $h > l$ $\frac{\psi(y^t|f_D^h)}{\psi(y^b|f_D^h)} > \frac{\psi(y^t|f_D^l)}{\psi(y^b|f_D^l)}$ (this is the MLRP adapted to the case at hands). In this case, I recover the insights from Proposition 1 after denoting $\theta_g^\dagger(s^i, \theta^i|y)$ the social reputation of an individual i with status s^i and ability θ^i after receiving signal y.

Proposition B.1. *For all $g \in \{D, d\}$, all $\theta^i \in [\overline{\theta}, \underline{\theta}]$, and all $s^i \in \{0, 1\}$, there exists a unique $y^0(s^i, \theta^i, g) \in (\underline{y}, \overline{y})$ such that*

- $\theta_g^\dagger(s^i, \theta^i|y^0(s^i, \theta^i, g)) = \theta_g^\dagger(s^i, \theta^i)$;
- *For all $y > (<) y^0(s^i, \theta^i, D)$, $\theta_D^\dagger(s^i, \theta^i|y) > (<) \theta_D^\dagger(s^i, \theta^i)$;*
  For all $y > (<) y^0(s^i, \theta^i, d)$, $\theta_d^\dagger(s^i, \theta^i|y) < (>) \theta_d^*(s^i, \theta^i)$.

If there exists an uninformative signal y^u such that $\psi(y^u|f^h) = \psi(y^u|f^l)$ for all $h \neq l$, then $y^0(s^i, \theta^i, g) = y^u$.

Proof. The proof proceeds very much along the lines of the proof of Proposition 1. I first focus on the dominant group. Denote $\mu_g^j(\theta^i, y) = Pr(\widetilde{f}_g = f_g^j|\theta^i, y)$. Repeating the steps to prove the second point of Lemma B.1, it can be shown that for all $y^t > y^b$ there exists a unique $m^0(y^t, y^b)$ satisfying $1 \le m^0(y^t, y^b) < n$ such that $\mu_D^j(\theta^i, y^t) < (\le) \mu_D^j(\theta^i, y^b)$ if $j < (\le) m^0(y^t, y^b)$ and $\mu_D^j(\theta^i, y^t) > \mu_D^j(\theta^i, y^b)$ if $j > m^0(y^t, y^b)$. Again repeating the same steps as in the proof of Lemma B.1, this implies: $\theta_D^\dagger(s^i, \theta^i|y^t) > \theta_D^\dagger(s^i, \theta^i|y^b)$ for all $y^t > y^b$.
The next step is to show that $\theta_D^\dagger(s^i, \theta^i|\underline{y}) < \theta_D^\dagger(s^i, \theta^i)$. To do so, I first prove that $\frac{\psi(\underline{y}|f_D^h)}{\psi(\underline{y}|f_D^l)} < 1$ for all $h > l$. By way of contradiction, suppose $\frac{\psi(\underline{y}|f_D^h)}{\psi(\underline{y}|f_D^l)} \ge 1$. Given the "MLRP", we have $\frac{\psi(y|f_D^h)}{\psi(y|f_D^l)} > \frac{\psi(\underline{y}|f_D^h)}{\psi(\underline{y}|f_D^l)}$ for all $y > \underline{y}$. This means that $\psi(y|f_D^h) > \psi(y|f_D^l)$ and $\psi(y|f_D^h) \ge \psi(y|f_D^l)$. Integrating over all y, we obtain: $1 > 1$, a contradiction.
With this, we can show that there exists a unique α^0 satisfying $1 \le \alpha^0 < n$ such that $\mu_D^j(\theta^i, \underline{y}) > (\ge) \mu_D^j(\theta^i)$ if $j < (\le) \alpha^0$ and $\mu_D^j(\theta^i, \underline{y}) < \mu_D^j(\theta^i)$ if $j > \alpha^0$. Notice that $\mu_D^j(\theta^i, \underline{y}) > \mu_D^j(\theta^i) \iff \sum_k^n \pi_g^k \frac{\psi(\underline{y}|f_g^k)}{\psi(\underline{y}|f_g^j)} < 1$. We necessarily have $\mu_D^1(\theta^i, \underline{y}) > \mu_D^1(\theta^i)$ and $\mu_D^n(\theta^i, \underline{y}) < \mu_D^n(\theta^i)$. Now suppose that for some $h \in \{2, ..., n\}$, we have $\mu_D^h(\theta^i, \underline{y}) \ge \mu_D^h(\theta^i)$. Take $j < h$ and notice that

$$\begin{aligned}\sum_k^n \pi_g^k \frac{\psi(\underline{y}|f_g^k)}{\psi(\underline{y}|f_g^j)} &= \sum_k^n \pi_g^k \frac{\psi(\underline{y}|f_g^k)}{\psi(\underline{y}|f_g^h)} \frac{\psi(\underline{y}|f_g^h)}{\psi(\underline{y}|f_g^j)} \\ &= \underbrace{\frac{\psi(\underline{y}|f_g^h)}{\psi(\underline{y}|f_g^j)}}_{<1} \underbrace{\sum_k^n \pi_g^k \frac{\psi(\underline{y}|f_g^k)}{\psi(\underline{y}|f_g^h)}}_{\le 1} < 1\end{aligned}$$

So if for some $h \in \{2, ..., n\}$, we have $\mu_D^h(\theta^i, \underline{y}) \ge \mu_D^h(\theta^i)$, then $\mu_D^j(\theta^i, \underline{y}) > \mu_D^j(\theta^i)$ for $j < h$. Similarly, if for some $h \in \{2, ..., n-1\}$, we have $\mu_D^h(\theta^i, \underline{y}) < \mu_D^h(\theta^i)$, then $\mu_D^j(\theta^i, \underline{y}) < \mu_D^j(\theta^i)$ for all $j > h$. All these elements together prove the existence and uniqueness of α^0.

We can then apply the same steps as in the proof of Lemma B.1 to establish that $\theta^{\dagger}_D(s^i,\theta^i|\underline{y}) < \theta^{\dagger}_D(s^i,\theta^i)$. Repeating the reasoning (and appropriately changing inequalities), we also obtain that $\theta^{\dagger}_D(s^i,\theta^i|\overline{y}) > \theta^{\dagger}_D(s^i,\theta^i)$. We can then apply the theorem of intermediate values to prove existence and uniqueness of $y^0(s^i,\theta^i,g) \in (\underline{y},\overline{y})$. For the disadvantaged group, we know that a high superscript for the dominant group means a low superscript for the disadvantaged group and, hence, all results are reverse.
Finally, the last point of Proposition B.1 follows from the same reasoning as for the proof of the last point of Proposition 1. □

Learning the composition of the elite

In this subsection, I assume that individuals are uncertain about both the values of the threshold and the distributions of ability. They, however, learn the composition of the elite. Hence, compared to the baseline model, I have substituted knowledge of the distributions of ability (assumptions 3 above) with knowledge about the composition of the elite (assumptions 5 above).
When it comes to uncertainty about the distributions of ability, I again denote $\mathcal{F}_g$ the prior set of possible distributions pdf f_g of ability over $[\underline{\theta},\overline{\theta}]$ for group g and $\widetilde{f}_g$ the random variable over the possible realization of f_g. As before, I assume that the set $\mathcal{F}_g$ contains countably many elements: $\mathcal{F}_g = \{f^1_g, f^2_g, f^3_g, ...\}$. I denote the cardinality of $\mathcal{F}_g$ by n (note that we can approach the continuous case by taking n to infinity) and assume that the last element in $\mathcal{F}_g$ is f^n_g. I assume that distributions are ranked in the sense of strict first order stochastic dominance. That is, I order the distribution so that $f^k_g > f^j_g$ if and only if F^k_g strictly first order stochastically dominates F^j_g (I also sometimes focus on the superscripts of the pdfs/CDFs since it is equivalent). The prior distribution satisfies: $Pr(\widetilde{f}_g = f^j_g) = \pi^j_g$ for $g \in \{d, D\}$, with $\pi^j_g > 0$ for all $j \in \{1, ..., n\}$. All distributions in $\mathcal{F}_g$ satisfy the conditions in the main text (i.e., all pdfs are continuous).
When it comes to uncertainty about the threshold values, I denote $\mathcal{E}_g$ the set of values $\widetilde{E}_g$ can take. I assume $\mathcal{E}_g$ is countable and of cardinality m so that $\widetilde{E}_g \in \{E^1_g = \underline{E}_g, E^2_g, ..., E^m_g = \overline{E}_g\}$. The values are ranked so that $E^h_g > E^l_g$ for all $h > l$. The prior distribution is $Pr(\widetilde{E}_g = E^j_g) = \gamma^j_g$.
Denote $\mathcal{R}$ the set of possible realizations of the share of individuals from the dominant group in the elite with $\mathcal{R} = \{\rho^1, \rho^2, ...\}$. Further, for all $\rho^h \in \mathcal{R}$, denote $\mathcal{K}_D(\rho^h) = \{f^j_D \in \mathcal{F}_D, E^k_D \in \mathcal{E}_D : \mathbb{E}_\epsilon(1 - F^j_D(E^k_D - \widetilde{\epsilon})) = \frac{e\times\rho^h}{\alpha}\}$. I assume that the cardinality of $\mathcal{D}(\rho^h)$ is strictly higher than one for all ρ^h (note that this implies that $\mathcal{R}$ has cardinality less than $\frac{nm}{2}$). I further assume that the distributions in the disadvantaged group are such that the set $\mathcal{K}_d(\rho^h) = \{f^j_d \in \mathcal{F}, E^k_d \in \mathcal{E}_d : \mathbb{E}_\epsilon(1 - F^j_d(E^k_d - \widetilde{\epsilon})) = \frac{e\times(1-\rho^h)}{1-\alpha}\}$ has also cardinality more than one for all values of ρ^h.
Notice importantly that each element in $\mathcal{K}_g(\rho^h)$ can easily be ranked: if the threshold E^k_g is high, then f^j_g is also high (in the sense of the order I have defined above). This means that one group can always justify its high representation in the elite by a high threshold **and** a high deservedness (a distribution of ability with a high mean).
We can use this observation to redefine the sets as $\mathcal{K}_g(\rho^h) = \{k^1_g(\rho^h), k^2_g(\rho^h), ...\}$ (i.e., each $k^l_g(\rho^h)$ is a particular realization of f^j_g and E^k_g) with cardinality and higher index $c(\rho^h)$ such that the elements of the sets are ordered in the following way: $t > b$ implies $\int_{\widetilde{\epsilon}} \mathbb{E}(\widetilde{\theta}|, s^i, \widetilde{\epsilon}, k^t_g(\rho^h)) > \int_{\widetilde{\epsilon}} \mathbb{E}(\widetilde{\theta}|, s^i, \widetilde{\epsilon}, k^b_g(\rho^h))$ for all elements in $\mathcal{K}_g(\rho^h)$. We can then define $\widetilde{k}_g(\rho^h)$ as the random variable over the possible values in the set $\mathcal{K}_g(\rho^h)$. Denote $\mu^l_g(\rho^h; s^i, \theta^i) = Pr(\widetilde{k}_g(\rho^h) = k^l_g(\rho^h)|s^i, \theta^i)$, the belief that $k^l_g(\rho^h)$ is realized given an individual i's ability and social status.
Building on the previous subsection, the social reputation is:

$$\theta^{\ddagger}_g(\rho^h; s^i, \theta^i) = \sum_{l=1}^{c(\rho^h)} \int_{\widetilde{\epsilon}} \mathbb{E}(\widetilde{\theta}|s^i, \widetilde{\epsilon}, k^l_g(\rho^h)) d\Lambda(\widetilde{\epsilon}) \mu^l_g(\rho^h; s^i, \theta^i) \tag{B.3}$$

I am now ready to define a public signal $x \in [\underline{x}, \overline{x}]$ with conditional CDF and pdf $X(\cdot|k^l_D(\rho^h))$ and $\chi(\cdot|k^l_D(\rho^h))$ (for all possible $k^l_D(\rho^h)$ for all possible ρ^h). For each $\rho^h \in \mathcal{R}$, I assume that a form of MLRP property holds: for each $x' > x$ and each $t > b$, $\frac{\chi(x|k^t_D(\rho^h))}{\chi(x'|k^t_D(\rho^h))} > \frac{\chi(x|k^b_D(\rho^h))}{\chi(x'|k^b_D(\rho^h))}$. Notice that I define the property within each realization of the share of group-D individuals in the elite (i.e., for each ρ^h).

Under these conditions, we can rank information into good news and bad news for the dominant group just as in the main text. Notice the importance of two conditions: the uncertainty is such that it matters for social reputation (this is given by the ordering I have assumed: $t > b$ implies $\int_{\widetilde{\epsilon}} \mathbb{E}(\widetilde{\theta}|s^i, \widetilde{\epsilon}, k_g^t(\rho^h)) > \int_{\widetilde{\epsilon}} \mathbb{E}(\widetilde{\theta}|s^i, \widetilde{\epsilon}, k_g^b(\rho^h))$) and the signals are easily separated into good news and bad news (this is given by the amended MLRP). Notice, however, that the information for group D does not contain any information for group d. Indeed, the two groups are now separate. What matters is how each group can justify its own proportion within the elite. As such, I obtain a watered down version of Proposition 1.

Proposition B.2. *For all $\theta^i \in [\overline{\theta}, \underline{\theta}]$, all $s^i \in \{0, 1\}$, and all $\rho^h \in \mathcal{R}$, there exists a unique $x^0(\rho^h; s^i, \theta^i) \in (\underline{x}, \overline{x})$ such that*

- $\theta_D^\ddagger(s^i, \theta^i | x^0(\rho^h; s^i, \theta^i)) = \theta_g^\ddagger(s^i, \theta^i; \rho^h)$;
- *For all $x > (<) x^0(s^i, \theta^i, D)$, $\theta_D^\ddagger(\rho^h; s^i, \theta^i | x) > (<) \theta_D^\ddagger(\rho^h; s^i, \theta^i)$.*

If there exists an uninformative signal $x^u(\rho^h)$ such that $\xi(x^u(\rho^h)|k_D^b(\rho^h)) = \psi(x^u(\rho^h)|k_D^t(\rho^h))$ for all $b \neq t$, then $x^0(\rho^h; s^i, \theta^i) = x^u(\rho^h)$.

Proof. The proof follows the same steps as the proof of Proposition B.1, after appropriately changing the notation. It is, thus, omitted. □

B.2 Robustness of amended model results

In the amended model, I make three assumptions: (1) Individuals know the thresholds to join the elite E_d and E_D, (2) Individuals know their ability, (3) Individuals know the size of the elite. The key force behind the results in the main text is that changing the threshold to enter the elite affects differently the chances of belonging to the elite and the social reputation. This differential effect is unaffected by relaxing the first and third assumptions, though this introduces noise and makes computations more difficult. Here, I discuss how the results change when individuals do not have perfect information about their ability.

As in Appendix B.1, suppose that each citizen i does not observe her ability θ^i, but receives instead a signal η^i distributed according to CDF $P(\eta^i|\theta^i)$ and pdf $p(\eta^i|\theta^i)$. The signal could be fully informative (in which case, $P(\eta^i|\theta^i)$ is a degenerate distribution), completely uninformative (in which case, $p(\eta^i|\theta^i) = p(\eta^i|\theta^{i\prime})$ for all $\eta^i, \theta^i, \theta^{i\prime}$ in their relevant supports), or anything in between. Given her signal η^i, an individual forms a posterior $F(\cdot|\eta^i)$ about the distribution of their ability. If from the dominant group D, her expected payoff is then:

$$\mathbb{E}_\epsilon(1 - F(E_D + \Delta - \epsilon|\eta^i))(1 + \theta_D^*(1, \Delta)) + \mathbb{E}_\epsilon(F(E_D + \Delta - \epsilon|\eta^i))(0 + \theta_D^*(0, \Delta))$$

For a citizen i from the disadvantaged group, the expected payoff is:

$$\mathbb{E}_\epsilon(1 - F(E_d - \delta(\Delta) - \epsilon|\eta^i))(1 + \theta_d^*(1, \delta(\Delta))) + \mathbb{E}_\epsilon(F(E_d - \delta(\Delta) - \epsilon|\eta^i))(0 + \theta_d^*(0, \delta(\Delta)))$$

The effect of changing the thresholds for a citizen i from the dominant and disadvantaged group is, respectively:

$$\begin{aligned}
&\mathbb{E}_\epsilon(F(E_D - \epsilon|\eta^i) - F(E_D + \Delta - \epsilon|\eta^i))(1 + \theta_D^*(1, \Delta) - \theta_D^*(0, \Delta)) \\
&+ \mathbb{E}_\epsilon(1 - F(E_D - \epsilon|\eta^i))(\theta_D^*(1, \Delta) - \theta_D^*(1, 0)) + \mathbb{E}_\epsilon(F(E_D - \epsilon|\eta^i))(\theta_D^*(0, \Delta) - \theta_D^*(0, 0)) \qquad \text{(B.4)} \\
&\mathbb{E}_\epsilon(F(E_d - \epsilon|\eta^i) - F(E_d - \delta(\Delta) - \epsilon|\eta^i))(1 + \theta_d^*(1, \delta(\Delta)) - \theta_d^*(0, \delta(\Delta))) \\
&+ \mathbb{E}_\epsilon(1 - F(E_d - \epsilon|\eta^i))(\theta_d^*(1, \delta(\Delta)) - \theta_d^*(1, 0)) + \mathbb{E}_\epsilon(F(E_d - \epsilon|\eta^i))(\theta_d^*(0, \delta(\Delta)) - \theta_d^*(0, 0)) \qquad \text{(B.5)}
\end{aligned}$$

Suppose that η^i is completely uninformative (i.e., each citizen has no private knowledge of their ability), then it is direct that Equation B.4 and Equation B.5 do not depend on the individual's ability. In other words, all individuals have the same payoff pre and post-reform. As such, I obtain:

Remark B.1. *Suppose that $p(\eta^i|\theta^i) = p(\eta^i|\theta^{i\prime})$ for all $\eta^i, \theta^i, \theta^{i\prime}$ in their relevant supports, then all citizens from group $g \in \{D, d\}$ either jointly support or jointly oppose changes to the conditions of entries into the elite.*

To describe in greater details the effect of uncertainty about ability on citizens' evaluation of the policies analyzed in this paper, I turn to a special case of the model where I assume that θ^i is normally distributed with mean 0 (without loss of generality) and variance σ_θ^2 and the random luck shock ϵ^i is normally distributed with mean zero and variance σ_ϵ^2. I further assume that the signal η^i that each citizen i receives takes the form of $\eta^i = \theta^i + \nu^i$ with $\nu^i \sim \mathcal{N}(0, \sigma_\nu^2)$. This approach is helpful to easily characterize the informativeness of an individual's signal. Indeed, by the conjugate prior property of the Normal distribution, an individual i's posterior distribution after signal η^i is $\mathcal{N}(\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\nu^2}\eta^i, \frac{\sigma_\theta^2\sigma_\nu^2}{\sigma_\theta^2+\sigma_\nu^2})$. As such, σ_ν^2 captures how informative i's signal is. The model studied in the main text corresponds to $\sigma_\mu^2 \to 0$ (slightly abusing notation). The case described in Remark B.1 corresponds to $\sigma_\nu^2 \to \infty$. In what follows, I suppose that $0 < \sigma_\nu^2 < \infty$.

Under the assumptions of this special case, notice first that individuals with a very high signal ($\eta^i \to \infty$) and a very low signal ($\eta^i \to -\infty$) see no change in their probability of joining the elite when changes to the entry condition into the elite are introduced. They are, respectively, certain to become elite member and sure to remain out of the elite. Those individuals always support policy reforms when they are from the dominant group (they benefit from the boost in social reputation) and always oppose quotas when they are from the disadvantaged group (they are hurt by the reputational loss). Individuals with signals close to the extremes see little changes in their chances of joining the elite due to the introduction of quotas and have the same perspective as those with infinitely high signals. So, as in the main text, only those who receive intermediary signals may have a different opinion about modifying the thresholds to ender the elite than individuals from their group with very large signals in absolute values. The question is can this division within group occurs when there is uncertainty about ability.

Proposition B.3 shows that the answer is yes when (i) luck does not play a very high part in an individual's success (in the formal language of the proposition, the variance of the luck shock must not be too large: σ_ϵ is strictly less than some threshold $\overline{\sigma}_\epsilon$) and (ii) the information individuals have about their ability cannot be too imprecise (in the formal language of the proposition, the variance of the signal σ_ν is strictly less than some threshold $\overline{\sigma}_\nu(\sigma_\epsilon)$).[3] This result is relatively intuitive, though the proof proves relatively complex. When luck plays a large role in success (i.e., its variance is large) and/or individuals know little about their ability (i.e., the signal is very imprecise), a small change in the threshold to join the elite will have little effect on individuals' evaluations of their chances of becoming an elite member. As such, they mostly consider the change in their social reputation, which goes in the same direction no matter their social status. Hence, in a setting with luck being much important and citizens not knowing much about their own ability, all members of the dominant group are likely to approve of a change to the thresholds for joining the elite and all members of the disadvantaged group rejects it. In contrast, when luck is not too important and citizens' knowledge of themselves is not too imprecise, then we recover a split within each group with the ends against the middle. As such, the result in the main text (Proposition 2) does not require individuals to know their ability, but still hold when the uncertainty about their own θ^i is not too large, at least for the special case of normally distributed ability, shock, and signals.

Proposition B.3. *There exist $\overline{\sigma}_\epsilon$ such that if $\sigma_\epsilon < \overline{\sigma}_\epsilon$, there exists $\overline{\sigma}_\nu(\sigma_\epsilon) > 0$ such that there exist unique finite η_D^l, η_D^h satisfying $\frac{\partial W_D(\eta^i,\Delta)}{\partial \Delta} < 0$ for all $\eta^i \in (\eta_D^l, \eta_D^h)$ and $\frac{\partial W_D(\eta^i,\Delta)}{\partial \Delta} \geq 0$ for all $\eta^i \notin (\eta_D^l, \eta_D^h)$ if and only if $\sigma_\nu < \overline{\sigma}_\nu(\epsilon)$.*
There exist $\widehat{\sigma}_\epsilon$ such that if $\sigma_\epsilon < \widehat{\sigma}_\epsilon$, there exists $\widehat{\sigma}_\nu(\sigma_\epsilon) > 0$ such that there exist unique finite η_d^l, η_d^h satisfying $\frac{\partial W_d(\eta^i,\delta)}{\partial \delta} > 0$ for all $\eta^i \in (\eta_d^l, \eta_d^h)$ and $\frac{\partial W_D(\eta^i,\delta)}{\partial \delta} \leq 0$ for all $\eta^i \notin (\eta_d^l, \eta_d^h)$ if and only if $\sigma_\nu < \widehat{\sigma}_\nu(\epsilon)$.

Proof. Consider an individual from the dominant group D with signal η^i. Notice that given the properties of the normal distribution, we obtain that $\theta^i + \epsilon^i | \eta^i \sim \mathcal{N}(\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\nu^2}\eta^i, \frac{\sigma_\theta^2\sigma_\nu^2}{\sigma_\theta^2+\sigma_\nu^2} + \sigma_\epsilon^2)$. As it is common, I use $\Phi(\cdot)$ and $\phi(\cdot)$ to denote respectively the CDF and pdf of the standard normal distribution. Denote $V^2 = \frac{\sigma_\theta^2\sigma_\nu^2}{\sigma_\theta^2+\sigma_\nu^2} + \sigma_\epsilon^2$, the

[3]When the first condition fails, it is possible that we end up in one of the extreme cases detailed in the proof of Proposition 2 even when ability is known.

expected payoff of this individual is:

$$W_D(\eta^i,\Delta) = \left(1-\Phi\left(\frac{E_D+\Delta-\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\nu^2}\eta^i}{V}\right)\right)\left(1+\theta_D^*(1,\Delta)\right)+\Phi\left(\frac{E_D+\Delta-\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\nu^2}\eta^i}{V}\right)\left(0+\theta_D^*(0,\Delta)\right) \tag{B.6}$$

Taking the derivative with respect to Δ, I obtain:

$$\frac{\partial W_D(\eta^i,\Delta)}{\partial\Delta} = -\frac{1}{V}\phi\left(\frac{E_D+\Delta-\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\nu^2}\eta^i}{V}\right)\left(1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)\right)+\left(1-\Phi\left(\frac{E_D+\Delta-\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\nu^2}\eta^i}{V}\right)\right)\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}+\Phi\left(\frac{E_D+\Delta-\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\nu^2}\eta^i}{V}\right)\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}$$

Observe that as $\eta^i\to\infty$, we obtain $\frac{\partial W_D(\eta^i,\Delta)}{\partial\Delta}>0$ (since $\phi\left(\frac{E_D+\Delta-\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\nu^2}\eta^i}{V}\right)\to 0$ and $\Phi\left(\frac{E_D+\Delta-\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\nu^2}\eta^i}{V}\right)\to 0$). Similarly, as $\eta^i\to-\infty$, we obtain $\frac{\partial W_D(\eta^i,\Delta)}{\partial\Delta}>0$ (since $\phi\left(\frac{E_D+\Delta-\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\nu^2}\eta^i}{V}\right)\to 0$ and $\Phi\left(\frac{E_D+\Delta-\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\nu^2}\eta^i}{V}\right)\to 1$). This corresponds to the observation made in the text.

Now consider how the derivative of $W_D(\eta^i,\Delta)$ wrt to Δ varies with signal η^i:

$$\begin{aligned}\frac{\partial^2 W_D(\eta^i,\Delta)}{\partial\Delta\partial\eta^i} =& \frac{\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\nu^2}}{V^2}\phi'\left(\frac{E_D+\Delta-\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\nu^2}\eta^i}{V}\right)\left(1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)\right)\\&+\frac{\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\nu^2}}{V}\phi\left(\frac{E_D+\Delta-\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\nu^2}\eta^i}{V}\right)\left(\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}\right)\end{aligned}$$

Using the properties of the normal distribution ($\phi'(x)=-x\phi(x)$), I obtain after rearranging that $\frac{\partial^2 W_D(\eta^i,\Delta)}{\partial\Delta\partial\eta^i}$ has the same sign as:

$$-\frac{E_D+\Delta-\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\nu^2}\eta^i}{V^2}\left(1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)\right)+\left(\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}\right)$$

Since the equation above is linear and strictly increasing in η^i, it is clear that there exists a unique $\eta^0(\sigma_\nu^2)$ so that $\frac{\partial^2 W_D(\eta^i,\Delta)}{\partial\Delta\partial\eta^i}$ is strictly negative (positive) for all $\eta^i<(>)\eta^0(\sigma_\nu^2)$.

Based on this observation, $\frac{\partial W_D(\eta^i,\Delta)}{\partial\Delta}$ reaches a minimum at $\eta^i=\eta^0(\sigma_\nu^2)$. Further, we have: $\frac{E_D+\Delta-\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\nu^2}\eta^0}{V} = V\frac{\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}}{1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)}$. Hence,

$$\begin{aligned}\frac{\partial W_D(\eta^0,\Delta)}{\partial\Delta} = \mathcal{W}_D(V) =& -\frac{1}{V}\phi\left(V\frac{\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}}{1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)}\right)\left(1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)\right)\\&+\left(1-\Phi\left(V\frac{\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}}{1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)}\right)\right)\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}\\&+\Phi\left(V\frac{\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}}{1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)}\right)\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}\end{aligned} \tag{B.7}$$

We then have (the derivative with respect to V should be understood as varying σ_ν^2)

$$\begin{aligned}
\mathcal{W}_D'(V) =&\frac{1}{V^2}\phi\left(V\frac{\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}}{1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)}\right)(1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta))\\
&-\frac{\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}}{V}\phi'\left(V\frac{\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}}{1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)}\right)\\
&-\frac{\left(\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}\right)^2}{1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)}\phi\left(V\frac{\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}}{1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)}\right)\\
\mathcal{W}_D'(V) =&\frac{1}{V^2}\phi\left(V\frac{\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}}{1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)}\right)(1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta))\\
&+\frac{\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}}{V}\times V\frac{\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}}{1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)}\phi\left(V\frac{\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}}{1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)}\right)\\
&-\frac{\left(\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}\right)^2}{1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)}\phi\left(V\frac{\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}}{1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)}\right)\\
=&\frac{1}{V^2}\phi\left(V\frac{\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}}{1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)}\right)(1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta))>0
\end{aligned}$$

Recall that $V^2=\frac{\sigma_\theta^2\sigma_\nu^2}{\sigma_\theta^2+\sigma_\nu^2}+\sigma_\epsilon^2$, so the lowest value V can take as we vary the informativeness of the signal is $V=\sigma_\epsilon$. Using Equation B.7, after noting that $\lim_{\sigma_\epsilon\to 0}\mathcal{W}_D(\sigma_\epsilon)=-\infty$ and $\lim_{\sigma_\epsilon\to\infty}\mathcal{W}_D(\sigma_\epsilon)>0$ ($\frac{1}{\sigma_\epsilon}\phi\left(\sigma_\epsilon\frac{\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}}{1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)}\right)$ goes to 0 as σ_ϵ goes to infinity and $\Phi\left(\sigma_\epsilon\frac{\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}}{1+\theta_D^*(1,\Delta)-\theta_D^*(0,\Delta)}\right)$ goes to zero or one depending on the sign of $\frac{\partial\theta_D^*(1,\Delta)}{\partial\Delta}-\frac{\partial\theta_D^*(0,\Delta)}{\partial\Delta}$). Hence, there exists a $\overline{\sigma}_\epsilon$ such that $\mathcal{W}_D(\sigma_\epsilon)$ is strictly positive (negative) whenever $\sigma_\epsilon<\overline{\sigma}_\epsilon$. Combining this with the properties of $\frac{\partial^2 W_D(\eta^i,\Delta)}{\partial\Delta\partial\eta^i}$ (strictly negative (positive) for all $\eta^i<(>)\eta^0(\sigma_\nu^2)$) and $\mathcal{W}_D(V)$, we obtain that if $\sigma_\epsilon<\overline{\sigma}_\epsilon$, then there exists a unique $\overline{\sigma}_\nu(\sigma_\epsilon)$ so that:

1. If $\sigma_\nu\geq\overline{\sigma}_\nu(\sigma_\epsilon)$, such that $\frac{\partial W_D(\eta^i,\Delta)}{\partial\Delta}>0$ for all η^i,
2. If $\sigma_\nu<\overline{\sigma}_\nu(\sigma_\epsilon)$, then there exists $\eta_D^l,\eta_D^h\in(-\infty,\infty)^2$ with $\eta_D^l<\eta_D^h$ such that $\frac{\partial W_D(\eta^i,\Delta)}{\partial\Delta}<0$ for all $\eta^i\in(\eta_D^l,\eta_D^h)$ and $\frac{\partial W_D(\eta^i,\Delta)}{\partial\Delta}\geq 0$ for all $\eta^i\notin(\eta_D^l,\eta_D^h)$.

A similar reasoning applies for members of the dominated group. □

C Additional formal results

C.1 Endogenous messages

In this appendix, I return to the baseline model with uncertainty about the thresholds . I suppose that the signal individuals receive is not exogenous, but consists of a message sends by a possibly strategic sender. I am interested in comparing how individuals react to messages coming from senders who share their group identity (in-group senders) and senders who come from the opposite group (out-group senders).

As noted in the main text, I build on Alonso and Padro i Miquel (2023) and I assume that individuals receive a message $m\in[\underline{z},\overline{z}]$ sent by an individual from group $g\in\{d,D\}$ who can either be honest (type $\tau=H$) or biased (type $\tau=B$). A honest sender observes z and always discloses it: $m(z)=z$. A biased sender does not observe z and only seeks to maximize the average social reputation of non-elite members from his group. Denote $\theta_g^*(s^i,\theta^i|m,G)$ the social reputation of an individual i from group $g\in\{d,D\}$, with social status $s^i\in\{0,1\}$ and ability θ^i conditional on receiving message $m\in[\underline{z},\overline{z}]$ from a sender from group $G\in\{d,D\}$ and $\theta_g^*(s^i|m,G)$ the associated average social reputation. A biased sender's payoff is equal to: $\theta_g^*(0|m,g)$.

The type of the sender is his private information and I assume that there is a probability π that the sender is honest. The public signal z has the same property as in the main text. It is distributed over the interval $[\underline{z},\overline{z}]$ with CDF and associated pdf conditional on the E_D (the realized threshold for group D): $Z(\cdot|E_D)$ and $\zeta(\cdot|E_D)$, respectively. The conditional distributions satisfy the strict monotone likelihood ration property:

$\frac{\zeta(z|E_D^h)}{\zeta(z'|E_D^h)} > \frac{\zeta(z|E_D^l)}{\zeta(z'|E_D^l)}$ for all $z > z'$, $E_D^h > E_D^l$. To facilitate the exposition, I assume that the distribution $\zeta(\cdot|E)$ contains a an uninformative signal z^u such that for all $E_D \neq E'_D$, $\zeta(z^u|E_D) = \zeta(z^u|E'_D)$.

The extended game proceeds as follows: Nature determines the realization of all random variables: each individual's ability θ^i, each citizen's luck ϵ^i, the entry thresholds into the elite E_D and E_d, the public signal z, and the type of the sender $\tau \in \{B, H\}$. The sender observes z if $\tau = H$ and nothing otherwise. The sender sends a message $m \in [\underline{z}, \overline{z}]$. Citizens in each group $g \in \{D, d\}$ with $\theta^i + \epsilon^i$ above the threshold E_g become elite members. Individuals observe the message m, their ability, and their social status, and compute their social reputation. Payoffs are realized.

Before proceeding to the analysis, let me explain why a biased sender maximizes the average social reputation of the non-elite members of his own group. To establish the strategies of a biased sender, it is helpful that this sender targets only one particular social reputation. I assume it is the average social reputation of the non-elite members since it seems in line with recent political events, but it could have been the elite members instead. As we will see, this also helps all individuals with the same group identity. As a result, this assumption on the objective of the biased sender appears to be without loss of generality.

To gain intuition on this extended model, let's consider individuals from the dominant group. First, let's assume that the sender is also from the dominant group. Using the notation introduces above, if the sender was known to be honest, then $\theta_D^*(s^i, \theta^i|m(z), D) = \theta_D^*(s^i, \theta^i|z)$, just like in the main text. If the sender is known to be biased, then the message is obviously completely uninformative and $\theta_D^*(s^i, \theta^i|m(z), D) = \theta_D^*(s^i, \theta^i)$, the expected social reputation absent any additional information. When there is uncertainty about the type, as Alonso and Padro i Miquel (2023) show, the biased sender can influence beliefs and, therefore, social reputation. Building on Alonso and Padro i Miquel (2023), I describe an equilibrium in which a biased sender only sends messages satisfying $m \geq z_D^B$.

First, note that all messages $m \geq z_D^B$ must induce the same average social reputation. Suppose not and there exists $m \geq z_D^B$ that maximizes the average social reputation of non-elite members from the dominant group (recall that this is the target audience of the biased sender by assumption). That is, $\theta_D^*(0|m, D) > \theta_D^*(0|m', D)$ for all $m' \neq m$. Then, the biased sender would only send message m, which would yield $\theta_D^*(0|m, D) < \theta_D^*(0|\hat{m}, D)$ for some $\hat{m}$ close enough to m, a contradiction. Further, if there exists one message m' such that $\theta_D^*(0|m, D) > \theta_D^*(0|m', D)$ for all $m \neq m'$, then the sender would never send message m' and the expected social reputation associated with m' would satisfy $\theta_D^*(0|m, D) < \theta_D^*(0|m', D)$, a contradiction. If all messages sent by a biased sender yield the same average social reputation, we must have $\theta_D^*(0|m, D) = \theta_D^*(0|z_D^B, D)$ for all $m \geq z_D^B$. Second, it must be that the biased sender prefers to send a message $m \geq z_D^B$ to any message $m < z_D^B$. Notice that given the strategy of the biased sender, any message $m' < z_D^B$ yields reputation $\theta_D^*(0|m', D)$ since the individuals believe that it can only be sent by a honest sender. If there exists m' such that $\theta_D^*(0|m', D) > \theta_D^*(0|z_D^B, D)$, the biased sender would deviate to message m', a contradiction.

Based on these observations, we can define the threshold z_D^B of the biased sender's strategy. Importantly, since there is always the possibility that a message $m > z_D^B$ is sent by a honest sender, the threshold z_D^B satisfies: $\theta_D^*(0|z_D^B, D) > \theta_D^*(0) \Leftrightarrow z_D^B > z^u$ (recall from Proposition 1 that for all $z > z^u$, social reputation increases for all individuals from the dominant group). In other words, individuals update positively on their expected social reputation after a high message even though they know that this high message may be sent by a biased sender. On the other end, if individuals from the dominant group receive a low message from an in-group sender (i.e., $m < z^u$), they update very negatively.

The analysis is quite obviously reversed for a sender from the disadvantaged group. A biased sender from the disadvantaged group sends message $m \leq z_d^B$ with the threshold z_d^B satisfying $z_d^B < z^u$. As such, members of the dominant group update slightly negatively after observing message $m \leq z_d^B$ since they take into account that such message can be sent by a biased sender. In turn, they would update very positively for all messages $m > z^u$ since they rightly understand that only an unbiased sender from the disadvantaged group sends such message.

As such, the analysis of this section reveals a few patterns. Messages matter for social reputation even if individuals rightly anticipate that some messages should be taken with a dose of skepticism. Second, how individuals update following a message depends on the identity of the sender, exactly because of this healthy skepticism. Fixing a message m, the expected social reputation of individuals from the dominant (disadvantaged) group is always weakly lower (weakly higher) after m if the sender is from the dominant group rather than the disadvantaged group. This plays a particular role for low message ($m < z_d^B$). In this case, the dominant group individuals would only update slightly negatively about their social reputation when the sender is from the disadvantaged group because there is a risk (or a hope) the sender is biased. However, their social reputation decreases massively after the same message from a sender of the dominant group as such message can only come from a honest sender. This extension, therefore, indicates that negative messages (in the sense that $m < z_d^B$) are to be expected from individuals of the disadvantaged group, but are perceived as a form of treason when they come from members of the dominant group.

The next proposition summarizes the findings of this section. To state it (and prove it), recall that I denote $\theta_g^*(s^i, \theta^i|m, G)$ the social reputation of an individual i from group $g \in \{d, D\}$, social status $s^i \in \{0,1\}$ and ability θ^i conditional on receiving message $m \in [\underline{z}, \overline{z}]$ from a sender from group $G \in \{d, D\}$. From the main text, recall that $\theta_g^*(s^i, \theta^i|z)$ is the social reputation when the signal is known to be z and $\theta_g^*(s^i, \theta^i)$ is the expected social reputation absent any additional information.

Proposition C.1. *There exist unique z_d^B, z_D^B satisfying $\underline{z} < z_d^B < z^u < z_D^B < \overline{z}$ such that:*

- *For all $m > z_D^B$, $\theta_D^*(s^i, \theta^i) < \theta_D^*(s^i, \theta^i|m, D) < \theta_D^*(s^i, \theta^i|m, d) = \theta_D^*(s^i, \theta^i|m)$ and $\theta_d^*(s^i, \theta^i) > \theta_d^*(s^i, \theta^i|m, D) > \theta_d^*(s^i, \theta^i|m, d) = \theta_d^*(s^i, \theta^i|m)$.*
- *For all $m < z_d^B$, $\theta_D^*(s^i, \theta^i) > \theta_D^*(s^i, \theta^i|m, d) > \theta_D^*(s^i, \theta^i|m, D) = \theta_D^*(s^i, \theta^i|m)$ and $\theta_d^*(s^i, \theta^i) < \theta_d^*(s^i, \theta^i|m, d) < \theta_d^*(s^i, \theta^i|m, D) = \theta_d^*(s^i, \theta^i|m)$.*
- *For $m \in [z_d^B, z_D^B]$, $\theta_g^*(s^i, \theta^i|m, D) = \theta_g^*(s^i, \theta^i|m, d) = \theta_g^*(s^i, \theta^i|m)$ for $g \in \{d, D\}$.*

Proof. To state the proof, it is useful to add some additional pieces of notation. Let $\theta_g^*(s^i, \theta^i|m, G, \tau)$ be the expected social reputation of an individual i from group $g \in \{d, D\}$, social status $s^i \in \{0,1\}$, type θ^i conditional on receiving message $m \in [\underline{z}, \overline{z}]$ from a sender from group $G \in \{d, D\}$ whose type is known to be $\tau \in \{H, B\}$. Obviously, $\theta_g^*(s^i, \theta^i|m, G, H) = \theta_g^*(s^i, \theta^i|m)$ and $\theta_g^*(s^i, \theta^i|m, G, B) = \theta_g^*(s^i, \theta^i)$. Further, let $Z(z)$ be the unconditional CDF of z and $\zeta(z)$ its associated pdf. Finally, let $\rho_G^B(m)$ be the pdf of the distribution of messages m is sent by a biased sender from group G.
To find the thresholds and their properties, I follow quite closely the proof of Proposition 1 in Alonso and Padro i Miquel (2023). There are a few differences worth stressing nonetheless. First, Alonso and Padro i Miquel (2023) consider biased senders who want to affect the posterior about a state of the world. In turn, I suppose that a biased sender from group G wants to maximize the average social reputation of non-elite members from his own group. Given the nature of the signal, this is equivalent to influence beliefs about z. Second, in Alonso and Padro i Miquel (2023), the receiver does not know whether the sender is biased in favour of one or the other state of the world. Here, I assume that the sender is biased in favour of its own group. This is without loss of generality since biased senders always send different signals in Alonso and Padro i Miquel (2023).
Consider a sender from group D. After observing message m, the average social reputation of individuals with status $s \in \{0,1\}$ takes value:

$$
\begin{aligned}
\theta_g^*(s|m, D) =& \frac{\pi\zeta(m)}{\pi\zeta(m) + (1-\pi)\rho_D^B(m)}\theta_g^*(s|m, D, H) + \frac{(1-\pi)\rho_D^B(m)}{\pi\zeta(m) + (1-\pi)\rho_D^B(m)}\theta_g^*(s|m, D, B) \\
=& \frac{\pi\zeta(m)}{\pi\zeta(m) + (1-\pi)\rho_D^B(m)}\theta_g^*(s|m) + \frac{(1-\pi)\rho_D^B(m)}{\pi\zeta(m) + (1-\pi)\rho_D^B(m)}\theta_g^*(s) \qquad \text{(C.1)}
\end{aligned}
$$

From Equation C.1, it can be seen that if m is such that $\theta_D^*(0|m, D) > \theta_D^*(0|m', D)$ for all $m' \neq m$, the biased sender's strategy is degenerate and always sends message m so that $\theta_D^*(0|m, D) = \frac{\pi\zeta(m)}{\pi\zeta(m)+(1-\pi)}\theta_D^*(0|m) + \frac{(1-\pi)}{\pi\zeta(m)+(1-\pi)}\theta_D^*(0)$. For any other m', $\theta_D^*(0|m', D) = \theta_D^*(0|m')$. It is immediate that for m' close enough to m if $m > z^u$ or for any $m > z^u$ if $m < z^u$, we have $\theta_D^*(0|m', D) > \theta_D^*(0|m, D)$, a contradiction. Notice that this directly implies $z_D^B < \overline{z}$. A similar reasoning explains why a biased sender's support contains all messages satisfying $m \geq z_D^B$ and why $\theta_D^*(0|m, D) = \theta_D^*(0|m', D) = \theta_D^*(0|z_D^B, D)$ for all $m, m' \geq z_D^B$, with a similar equality holding for the disadvantaged group.
As a result, for all $m \geq z_D^B$, we obtain from Equation C.1

$$
\begin{aligned}
\theta_D^*(0|z_D^B, D) =& \frac{\pi\zeta(m)}{\pi\zeta(m) + (1-\pi)\rho_D^B(m)}\theta_D^*(0|m) + \frac{(1-\pi)\rho_D^B(m)}{\pi\zeta(m) + (1-\pi)\rho_D^B(m)}\theta_D^*(0) \\
\Leftrightarrow (1-\pi)\rho_D^B(m)(\theta_D^*(0|z_D^B, D) - \theta_D^*(0)) =& \pi\zeta(m)(\theta_D^*(0|m) - \theta_D^*(0|z_D^B, D)) \qquad \text{(C.2)}
\end{aligned}
$$

Integrating Equation C.2 for all $m \geq z_D^B$, I obtain:

$$(1-\pi)(\theta_D^*(0|z_D^B, D) - \theta_D^*(0)) = \pi \int_{z_D^B}^{\overline{z}} \theta_D^*(0|m) - \theta_D^*(0|z_D^B, D)dZ(m) \tag{C.3}$$

Equation C.3 determines the unique z_D^B (using the same steps as in Proposition 1, it can be shown that $\theta_D^*(s^i|z)$ is strictly increasing with z and so is $\theta_D^*(0|z, D)$ by Equation C.1). Notice further that $z_D^B > z^u$ (otherwise the left-hand side is zero and the right-hand side is strictly positive).

Now given the properties of social reputation, the average social reputation of individuals from group g with status $s \in \{0,1\}$ is: $\theta_g^*(s|z_D^B, D)$ for all $m \geq z_D^B$ and $\theta_g^*(s|m, D) = \theta_g^*(s|m)$ for all $m < z_D^B$, with z_D^B defined by Equation C.3.

Taking a sender from the disadvantaged group and applying the same reasoning, I obtain that a biased sender sends message $m \leq z_d^B \in (\underline{z}, z^u)$ with the threshold defined by $(1-\pi)(\theta_d^*(0|z_d^B, d) - \theta_d^*(0)) = \pi \int_{\underline{z}}^{z_d^B} \theta_d^*(0|m) - \theta_d^*(0|z_d^B, d)dZ(m)$. As a result, the average social reputation of individuals from group g with status $s \in \{0,1\}$ is: $\theta_g^*(s|z_d^B, d)$ for all $m \leq z_d^B$ and $\theta_g^*(s|m, d) = \theta_g^*(s|m)$ for all $m > z_d^B$.

Given the relationship between the social reputations of the two groups (see the proof of Proposition 1) and Equation C.2, we necessarily have for all $s \in \{0,1\}$:

- For all $m > z_D^B$, $\theta_D^*(s) < \theta_D^*(s|m, D) < \theta_D^*(s|m, d) = \theta_D^*(s|m)$ and $\theta_d^*(s) > \theta_d^*(s|m, D) > \theta_d^*(s|m, d) = \theta_d^*(s|m)$.
- For all $m < z_d^B$, $\theta_D^*(s) > \theta_D^*(s|m, d) > \theta_D^*(s|m, D) = \theta_D^*(s|m)$ and $\theta_d^*(s) < \theta_d^*(s|m, d) < \theta_d^*(s|m, D) = \theta_d^*(s|m)$.
- For $m \in [z_d^B, z_D^B]$, $\theta_g^*(s|m, D) = \theta_g^*(s|m, d) = \theta_g^*(s|m)$ for $g \in \{d, D\}$.

To finish the proof, note that for an individual from group $g \in \{d, D\}$ with ability θ^i and status s^i, we can write the social reputation after message m from a sender from the dominant group as:

$$\begin{aligned}
\theta_g^*(s^i, \theta^i|m, D) =& \frac{\pi\zeta(m)}{\pi\zeta(m) + (1-\pi)\rho_D^B(m)}\theta_g^*(s^i, \theta^i|m) + \frac{(1-\pi)\rho_D^B(m)}{\pi\zeta(m) + (1-\pi)\rho_D^B(m)}\theta_g^*(s^i, \theta^i) && \text{if } m \geq z_D^B \\
\theta_g^*(s^i, \theta^i|m, D) =& \theta_g^*(s^i, \theta^i|m) && \text{if } m < z_D^B
\end{aligned}$$

In turn, when the sender is from the disadvantaged group, the expected social reputation of the same individual after message m is:

$$\begin{aligned}
\theta_g^*(s^i, \theta^i|m, d) =& \frac{\pi\zeta(m)}{\pi\zeta(m) + (1-\pi)\rho_d^B(m)}\theta_g^*(s^i, \theta^i|m) + \frac{(1-\pi)\rho_d^B(m)}{\pi\zeta(m) + (1-\pi)\rho_d^B(m)}\theta_g^*(s^i, \theta^i) && \text{if } m \leq z_d^B \\
\theta_g^*(s^i, \theta^i|m, d) =& \theta_g^*(s^i, \theta^i|m) && \text{if } m > z_d^B
\end{aligned}$$

Since $z_D^B > z^u > z_d^B$, it must be that for all $m > z_D^B$, $\theta_D^*(s^i, \theta^i|m) > \theta_D^*(s^i, \theta^i)$ so that $\theta_D^*(s^i, \theta^i|m, d) = \theta_D^*(s^i, \theta^i|m) > \theta_D^*(s^i, \theta^i|m, D)$. In turn, for all $m < z_d^B$, then $\theta_D^*(s^i, \theta^i|m) < \theta_D^*(s^i, \theta^i)$ so that $\theta_D^*(s^i, \theta^i|m, D) = \theta_D^*(s^i, \theta^i|m) < \theta_D^*(s^i, \theta^i|m, d)$. For all $m \in (z_d^B, z_d^B)$, the expected social reputation satisfies: $\theta_D^*(s^i, \theta^i|m, d) = \theta_D^*(s^i, \theta^i|m, d) = \theta_D^*(s^i, \theta^i|m)$. By continuity, the equality must also be true at $m = z_D^B$ and $m = z_d^B$. Since we know that the social reputation of members of the disadvantaged group moves in the opposite direction, we obtain the result. □

C.2 Uncertainty about the distribution of abilities

In this last formal supplementary appendix, I sketch a simpler model with uncertainty about the distribution of abilities in the dominant group. I assume that ability in group $g \in \{d, D\}$ is uniformly distributed over the interval $[-\overline{\theta} + k_g, \overline{\theta} + k_g]$, with $\overline{\theta} > 1$. I further assume that while $k_d = 0$ is common knowledge, k_D is uncertain. However, it is commonly known that $k_D \in \{0, D\}$ with $0 < D < 1$ and $Pr(k_D = D) = \pi$. As such, the distribution of ability among the dominant group is either equal to that of the disadvantaged group or higher. There is, thus, a possibility that the dominant group is more deserving than the disadvantaged group.

Of course, this better distribution of skills may be due to past discrimination, but for my concern, I take it as given.

I make a few further assumptions to facilitate the analysis: (i) the distribution of the threshold $\widetilde{E}_D$ takes two values: $\widetilde{E}_D \in \{0, 1\}$ with $Pr(\widetilde{E}_D = 1) = \gamma$, (ii) individuals do not know their ability, and (iii) luck plays no role.[4] In turn, like in the main model, the size of the elite is known, whereas the threshold for the disadvantaged group is not.

Under the assumptions above, the proportion of dominant group members that make it into the elite can take one of four values:

- $P_1 = \frac{\bar{\theta}+D}{2\bar{\theta}}$ if $k_D = D$ and $E_D = 0$ (i.e., with probability $\pi(1-\gamma)$),
- $P_2 = \frac{1}{2}$ if $k_D = 0$ and $E_D = 0$ (i.e., with probability $(1-\pi)(1-\gamma)$),
- $P_3 = \frac{\bar{\theta}+D-1}{2\bar{\theta}}$ if $k_D = D$ and $E_D = 1$ (i.e., with probability $\pi\gamma$),
- $P_4 = \frac{\bar{\theta}-1}{2\bar{\theta}}$ if $k_D = 0$ and $E_D = 1$ (i.e., with probability $(1-\pi)\gamma$).

These proportions are ranked as: $P_1 > P_2 > P_3 > P_4$. Each proportion, you will notice, is associated with a different threshold for the disadvantaged group which I can rank as $E_d^1 > E_d^2 > E_d^3 > E_d^4$.

In this setting, we can think of two types of public signal that still maintain some uncertainty about the distribution of abilities in the dominant group. The first is a public message that reveals members from group D constitute strictly more than $\alpha P_4/e$. I call this signal z_1. The second message is that members from group D constitute strictly more than $\alpha P_3/e$ of the elite. I label this signal z_2.

Absent any information, individuals evaluate an elite member from group D based on the chances a high-ability individual makes it to the elite relative to a low-ability one across the four events above, which can broadly be summarized as high/low share of high-ability individuals, easy/hard threshold to reach to join the elite. With the first signal (z_1), everyone knows that it is not possible to have simultaneously a distribution of ability in the dominant group equal to the distribution in the disadvantaged group and a hard threshold for joining the elite. Hence, signal z_1 provides both good news (regarding the distribution of types) and bad news (regarding the threshold) for the social reputation of individuals from group D. Good news dominates when the gain from putting more weight on a better distribution of ability in the dominant group, which is proportional to πD , is higher than the loss from the higher chances of an easy threshold, which is proportional to $(1-\gamma) \times 1$. As such, uncertainty about abilities can serve as an "excuse" to actually improve the social reputation of the dominant group only if the disadvantaged group is viewed as sufficiently undeserving (in term of probability or differences in ability).

In turn, it is easy to see why signal z_2 necessarily hurts the social reputation of group-D members. After observing z_2, every citizen faces uncertainty about the distribution of types, but all know that the threshold for entering the elite for individuals from the dominant group is low ($\widetilde{E}_D = 0$). Hence, the social reputation of group-D citizens necessarily decrease relative to a setting with no information.

Notice that after information z_1 or z_2, the social reputation of individuals from the disadvantaged group necessarily increase. This signal indicates that more weight should be put on high thresholds than on low thresholds for the disadvantaged group. Here, we recover the first-order stochastic dominance effect at play in the main text.

Overall, the analysis of this section reveals that uncertainty about the proportion of high-ability type in the dominant group yields some interesting patterns. The possibility of explaining the dominant group success by its greater deservedness can help the social reputation of the dominant group, but not always. There are still cases where Proposition 1 holds and public information hurts all the individuals from the dominant group. Further, even when it helps the dominant group, the effect of information is the same for all members of the same group, regarding of their social status. While a full analysis is left for future research, the amended model presented here suggests that the results are not necessarily overturned by the introduction of second-order uncertainty.

The findings of this section are summarized in Proposition C.2. I denote $\theta_g^*(s^i)$ the (expected) social ability of a group-g individual with status s^i absent information (remember that individuals do not know their ability). In turn, $\theta_g^*(s^i|z)$ is the (expected) social ability after signal $z \in \{z_1, z_2\}$ (recall that z_1 states that group-D

[4] These last two assumptions do not play an important role in establishing Proposition 1. Here, they make the analysis much simpler.

individuals constitute strictly more than $\alpha P_4/e$ percent of the elite and z_2 that they constitute strictly more than $\alpha P_3/e$ percent of the elite).

Proposition C.2. *For the dominant group, $\theta^*_D(s^i) > (<)\theta^*_D(s^i|z_1)$ for all $s^i \in \{0,1\}$ if and only if $1-\gamma > (<)\pi D$. For all $s^i \in \{0,1\}$, $\theta^*_D(s^i) > \theta^*_D(s^i|z_2)$.*
For the disadvantaged group, $\theta^*_d(s^i) < \theta^*_d(s^i|z)$ for all $s^i \in \{0,1\}$ and $z \in \{z_1, z_2\}$.

Proof. Consider a member of the dominant group. Absent information, his social reputation is for elite and non-elite status, respectively:

$$\theta^*_D(1) = \overbrace{\frac{\overline{\theta}+D}{2}\pi(1-\gamma)}^{k_D=D,E_D=0} + \overbrace{\frac{\overline{\theta}}{2}(1-\pi)(1-\gamma)}^{k_D=0,E_D=0} + \overbrace{\frac{\overline{\theta}+D+1}{2}\pi\gamma}^{k_D=D,E_D=1} + \overbrace{\frac{\overline{\theta}+1}{2}(1-\pi)\gamma}^{k_D=0,E_D=1} \tag{C.4}$$

$$\theta^*_D(0) = \overbrace{\frac{-\overline{\theta}+D}{2}\pi(1-\gamma)}^{k_D=D,E_D=0} + \overbrace{\frac{-\overline{\theta}}{2}(1-\pi)(1-\gamma)}^{k_D=0,E_D=0} + \overbrace{\frac{-\overline{\theta}+D+1}{2}\pi\gamma}^{k_D=D,E_D=1} + \overbrace{\frac{-\overline{\theta}+1}{2}(1-\pi)\gamma}^{k_D=0,E_D=1} \tag{C.5}$$

After signal $z = z_1$, the social reputations are:

$$\theta^*_D(1|z_1) = \overbrace{\frac{\overline{\theta}+D}{2}\frac{\pi(1-\gamma)}{(1-\gamma)+\pi\gamma}}^{k_D=D,E_D=0} + \overbrace{\frac{\overline{\theta}}{2}\frac{(1-\pi)(1-\gamma)}{(1-\gamma)+\pi\gamma}}^{k_D=0,E_D=0} + \overbrace{\frac{\overline{\theta}+D+1}{2}\frac{\pi\gamma}{(1-\gamma)+\pi\gamma}}^{k_D=D,E_D=1} \tag{C.6}$$

$$\theta^*_D(0|z_1) = \overbrace{\frac{-\overline{\theta}+D}{2}\frac{\pi(1-\gamma)}{(1-\gamma)+\pi\gamma}}^{k_D=D,E_D=0} + \overbrace{\frac{-\overline{\theta}}{2}\frac{(1-\pi)(1-\gamma)}{(1-\gamma)+\pi\gamma}}^{k_D=0,E_D=0} + \overbrace{\frac{-\overline{\theta}+D+1}{2}\frac{\pi\gamma}{(1-\gamma)+\pi\gamma}}^{k_D=D,E_D=1} \tag{C.7}$$

Simple algebra yield the result.
In turn, for $z = z_2$, social reputations are:

$$\theta^*_D(1|z_2) = \overbrace{\frac{\overline{\theta}+D}{2}\pi}^{k_D=D,E_D=0} + \overbrace{\frac{\overline{\theta}}{2}(1-\pi)}^{k_D=0,E_D=0} \tag{C.8}$$

$$\theta^*_D(0|z_2) = \overbrace{\frac{-\overline{\theta}+D}{2}\pi}^{k_D=D,E_D=0} + \overbrace{\frac{-\overline{\theta}}{2}(1-\pi)}^{k_D=0,E_D=0} \tag{C.9}$$

Quite clearly, the claim holds.
The result for the disadvantaged group follows from the observation that signals z_1 and z_2 lead to more weight being put on the more stringent thresholds relative to the case without information.

□

D Empirical analysis

In this section, I present the results of the empirical analysis of the British Election Study, General Social Survey, and Cooperative Election Study. Information on the dependent variables used can be found in the notes of the table. Regressions with controls include variables on education (university or finished high school), home ownership, marital situation, age, income, working status, working sector, wave or year fixed effects, religion fixed effects, and (when possible) location fixed effects. All regressions are OLS regressions with robust standard errors. For more details on data sources, variable constructions, and empirical specifications, see the documentation for this article available on the APSR Dataverse at https://doi.org/10.7910/DVN/B3P41O.

D.1 Happiness

British Election Study

Table D.1 reports the result on self-rated happiness and life worthiness from the BES. Absent controls, white men are slightly more likely to report that they are happy (column (1)). Yet, white men tend to be more successful on average and success may bring happiness. When I include controls that proxy for social success (income, education, owning houses), the coefficient changes signs and becomes highly significant (column (2)). While the size of the coefficient is relatively small relative to the mean, the difference between white men and other respondents equals more than half the difference between renters and owners or is equal to the difference between divorcees and singles or in cohabitation (see Table F.1 in the document Angry White Males - Dataverse.pdf on the APSR Dataverse for this article). Notice that this is very much a white *male* phenomenon as when I restrict the sample to whites (column (3))—so that the reference category is white women—, the coefficients remain unchanged. When it comes to life worthiness, white men are always less likely to rate their life lower, with or without controls, when they are compared to all respondents or just white women (columns (4) to (6)).

Table D.1: Self-reported happiness and life worthiness in the UK (2014-2023)

	(1)	(2)	(3)	(4)	(5)	(6)
	Happy yesterday			Life worthwhile		
White Male	0.066*	-0.139***	-0.147***	-0.207***	-0.393***	-0.388***
	(0.054)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sample	All	All	White	All	All	White
Mean dep variable	6.07	6.09	6.11	6.22	6.23	6.24
Individual controls		✓	✓		✓	✓
N.obs	21954	20811	19280	21611	20484	19006

Notes: Dependent variables are categorical variable from 0 (not at all) to 10 (very). Complete model results can be found in Table F.1 in the Angry White Males - Dataverse.pdf file available in the APSR Dataverse for this article. Robust standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

General Social Survey

The patterns are the same when we look at US data from the General Social Survey. For ease of comparison with the UK data, Table D.2 restricts the sample to respondents interviewed in 2014 or after. Absent controls for social success, white men are moderately more happy than other respondents (column (1)). With controls, the coefficient on white male becomes highly significant and negative even after restricting the sample to whites (columns (2) and (3)). The magnitude of the effect is also similar: more than half the difference between renters and owners, equal to the differences between divorcees and singles or in cohabitation (see Table F.2 in the document Angry White Males - Dataverse.pdf on the APSR Dataverse for this article).

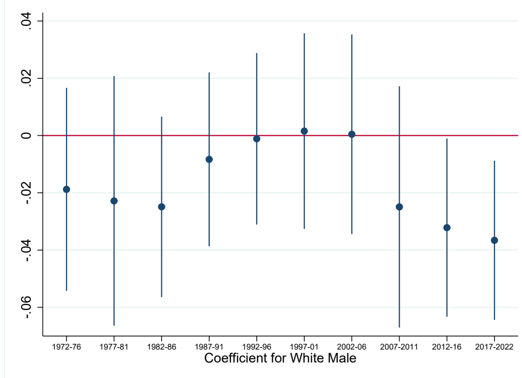
Table D.2: Self-reported happiness in the USA (2014-2022)

	(1)	(2)	(3)
	Self-rated happiness		
White Male	0.011	-0.040***	-0.036***
	(0.350)	(0.000)	(0.004)
Sample	All	All	White
Mean dep variable	1.07	1.07	1.08
Individual controls		✓	✓
N.obs	15267	14547	10825

Notes: Happy is a categorical variable from 0 (not too happy) to 2 (very). Complete model results can be found in Table F.2 in the Angry White Males - Dataverse.pdf file available in the APSR Dataverse for this article. Robust standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

I take advantage of the full GSS data and look at the evolution of white men's happiness relative to other respondents over time. To limit sample variations, I group surveys in 5-year periods from 1972 until the last available data (6-year for the last period, though the relevant question was not asked in 2017 and 2018). I

Figure D.1: Self-reported happiness over time in the USA



Notes: Happy is a categorical variable from 0 (not too happy) to 2 (very). Complete model results can be found in Table F.3 in the Angry White Males - Dataverse.pdf file available in the APSR Dataverse for this article. Dots represent point estimates and vertical lines display the 95% confidence intervals.

plot the coefficients on white men from the regression displayed in column (2) of Table D.2 (with all controls). Figure D.1 reveals three distinct periods. White men were on average less happy than other respondents until the end of the 1980s. They were as happy as women and minorities in the 1990s and until 2006. They returned to a lower level of reported happiness afterwards. Further, the difference between white men and other groups seem to be greater nowadays than at any point in time.

D.2 Additional results: Information vs Policy

British Election Study

Table D.3 looks at attitudes towards policies in favour of minorities (columns (1) and (2)), women (columns (3) and (4)), lesbian and gays (columns (5) and (6)). With or without controls, the findings are always the same. White men are more likely to oppose such policies. The effects are quite substantial between one fourth and 50% relative to the mean.

Table D.3: Attitudes on policies toward minorities

	(1)	(2)	(3)	(4)	(5)	(6)
Equal opport. to	Minorities gone too far		Women gone too far		Lesbians-Gays gone too far	
White Male	0.098^{***}	0.078^{***}	0.072^{***}	0.069^{***}	0.124^{***}	0.102^{***}
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sample	All	All	All	All	All	All
Mean dep variable	0.29	0.30	0.14	0.14	0.27	0.27
Individual controls		✓		✓		✓
N.obs	169545	162210	169761	162426	169545	162210

Notes: Dependent variables are indicator variables taking value 1 if respondent believes policies have gone too far or much too far. Complete model results can be found in Table F.4 in the Angry White Males - Dataverse.pdf file available in the APSR Dataverse for this article. Robust standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.4 looks at the differences between white men and other respondents on policies towards disadvantaged groups by level of education. As respondents' level of education increases, they become less likely to oppose improving equal opportunity for the various groups considered (see the row titled mean dep. variable). For every disadvantaged group, however, the difference in attitudes between white men and other respondents remains constant. White men are around 8% more likely to state that equal opportunities to minorities have gone too far, 7% more likely to state that equal opportunities to women have gone too far, and 10% more likely to state that equal opportunities to lesbians and gays have gone too far. These findings are much more aligned with white men's anger being triggered by information rather than by policy changes as noted in the main text.

Table D.5 looks at the differences between white men and other respondents on policies towards disadvantaged groups by age groups. Here again, we see little differences between age groups. One exception is policies

Table D.4: Attitudes on policies toward minorities by level of educations

Equal opport. to	(1)	(2) Minorities gone too far	(3)	(4)	(5) Women gone too far	(6)	(7)	(8) Lesbians-Gays gone too far	(9)
White Male	0.080***	0.075***	0.084***	0.061***	0.068***	0.072***	0.102***	0.112***	0.095***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sample	No qualif./answer	High School	University	No qualif./answer	High School	University	No qualif./answer	High School	University
Mean dep variable	0.36	0.35	0.24	0.14	0.15	0.12	0.32	0.30	0.23
Individual controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
N.obs	16100	72483	73627	16115	72580	73731	16100	72483	73627

Notes: Dependent variables are indicator variables taking value 1 if respondent believes policies have gone too far or much too far. Complete model results can be found in Table F.5 in the Angry White Males - Dataverse.pdf file available in the APSR Dataverse for this article. Robust standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

in favour of women with over-65 white men being closer to the attitudes of other groups than younger age groups (the highest difference between white men and others for this item is actually for under-25 respondents, consistent with the observed divergence on feminism, mentioned in the introduction). Yet the coefficient on white men for over 65 is only one third smaller than the coefficient for 26-64 years old. The evidence in favour of policy changes favouring disadvantaged groups as a source of white men's anger is, thus, limited.

Table D.5: Attitudes on policies toward minorities by age groups

Equal opport. to	(1)	(2) Minorities gone too far	(3)	(4)	(5) Women gone too far	(6)	(7)	(8) Lesbians-Gays gone too far	(9)
White Male	0.106***	0.068***	0.085***	0.096***	0.074***	0.051***	0.055***	0.084***	0.144***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sample	Under 25	26-64	Over 65	Under 25	26-64	Over 65	Under 25	26-64	Over 65
Mean dep variable	0.14	0.28	0.38	0.11	0.13	0.15	0.12	0.23	0.40
Individual controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
N.obs	11815	103471	46924	11829	103606	46991	11815	103471	46924

Notes: Dependent variables are indicator variables taking value 1 if respondent believes policies have gone too far or much too far. Complete model results can be found in Table F.6 in the Angry White Males - Dataverse.pdf file available in the APSR Dataverse for this article. Robust standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

When it comes to opinions about discrimination, white men are more likely to say that men or whites are discriminated and less likely to agree that women or ethnic minorities (BME) are discriminated as shown in Table D.6. Again, this holds with or without controls.

Table D.6: Attitudes on discrimination

	(1) Men discriminated	(2)	(3) Women discriminated	(4)	(5) White discriminated	(6)	(7) BME discriminated	(8)
White Male	1.413***	1.352***	-0.867***	-0.788***	0.652***	0.547***	-0.698***	-0.598***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sample	All	All	All	All	All	All	All	All
Mean dep variable	4.08	4.09	5.76	5.75	4.69	4.73	5.91	5.89
Individual controls		✓		✓		✓		✓
N.obs	77037	73834	78832	75560	77812	74616	78297	75072

Notes: Dependent variables are categorical variable from 0 (a lot of discrimination in favour) to 10 (a lot of discrimination against). Complete model results can be found in Table F.7 in the Angry White Males - Dataverse.pdf file available in the APSR Dataverse for this article. Robust standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In Table D.7, I look at how likely white men are to state that men and whites are discriminated relative to other groups by level of education. Again, we observe that education is associated with a lower propensity to state that whites or men are discriminated (see the mean of dep. variable row). Yet, the coefficient on white men is very similar across all columns and, if anything, it is higher for white men with university degree than others. As such, the evidence presented in Table D.7 are more consistent with white men's anger triggered by information than by policy changes.

Looking at opinions on discrimination against whites and men by age groups in Table D.8, quite strikingly, among the under-25, white men are much more likely to state that whites are discriminated, consistent with the finding that there is a growing liberal divide between men and other groups as noted in the introduction (column (4)). We also observe patterns more consistent with a policy effect (at least for discrimination against whites, columns (4) to (6)). Indeed, white men have less distinct attitudes than women and minorities in the

Table D.7: Attitudes on discrimination by level of educations

Equal opport. to	(1)	(2) Men discriminated	(3)	(4)	(5) White discriminated	(6)
White Male	1.103***	1.361***	1.371***	0.372***	0.482***	0.614***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sample	No qualif./answer	High School	University	No qualif./answer	High School	University
Mean dep variable	4.45	4.20	3.92	5.49	5.04	4.30
Individual controls	✓	✓	✓	✓	✓	✓
N.obs	5920	32698	35216	6149	33155	35312

Notes: Dependent variables are categorical variable from 0 (a lot of discrimination in favour) to 10 (a lot of discrimination against). Complete model results can be found in Table F.8 in the Angry White Males - Dataverse.pdf file available in the APSR Dataverse for this article. Robust standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

over-65 age group than in other groups. The coefficient in column (3), however, is only 20% smaller than the coefficient in column (2). In turn, the coefficient in column (6) is 32% smaller than the coefficient in column (5). Hence, even if policies matter, the results suggest there is still room for a substantively significant effect of information.

Table D.8: Attitudes on discrimination by age groups

	(1)	(2) Men discriminated	(3)	(4)	(5) White discriminated	(6)
White Male	1.427***	1.445***	1.156***	1.266***	0.563***	0.388***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sample	Under 25	26-64	Over65	Under 25	26-64	Over65
Mean dep variable	3.42	4.03	4.34	3.23	4.69	5.11
Individual controls	✓	✓	✓	✓	✓	✓
N.obs	4860	45642	23332	4846	46027	23743

Notes: Dependent variables are categorical variable from 0 (a lot of discrimination in favour) to 10 (a lot of discrimination against). Complete model results can be found in Table F.9 in the Angry White Males - Dataverse.pdf file available in the APSR Dataverse for this article. Robust standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

General Social Survey

I now turn to variables in the General Social Survey. To facilitate comparisons with the other surveys, and given the evolution over time noted in Figure D.1, I restrict the sample to the post-2014 surveys. Table D.9 considers opposition to affirmative action (columns (1) and (2)), beliefs that Blacks should find their way up without assistance (resentment item in columns (3) and (4)), and beliefs that there is too much spending on assistance to Blacks (columns (5) and (6)), too much is spent on the improvement of Blacks (columns (7) and (8)). In all cases, with or without controls, white men hold much less favourable views to policies that benefit African-Americans.

Table D.9: Policy attitudes towards Blacks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Oppose affirmative action		Resentment		Too much assistance		Too much on improvement	
White Male	0.058***	0.058***	0.042***	0.033***	0.030***	0.029***	0.058***	0.058***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sample	All	All	All	All	All	All	All	All
Mean dep variable	0.85	0.85	0.36	0.36	0.06	0.06	0.08	0.08
Individual controls		✓		✓		✓		✓
N.obs	15329	14596	15329	14596	15329	14596	15329	14596

Notes: Dependent variables are indicator variables. For columns (1) and (2), variable equals one if respondent opposes (strongly or not strongly) affirmative action, 0 otherwise. For columns (3) and (4), variable equals one if respondent agrees (somewhat or strongly) that Blacks should overcome prejudice without favors and 0 otherwise. Columns (5) to (8), variable equals one if respondent states that the US spends too much on improving the conditions of/on assistance to Blacks. Complete model results can be found in Table F.10 in the Angry White Males - Dataverse.pdf file available in the APSR Dataverse for this article. Robust standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In Table D.10, I look at attitudes on spending for the assistance (columns (1)-(3)) and for the improvement (columns (4)-(6)) of Blacks by levels of education. As for the British Election Survey, respondents with higher level of education are less likely to state that too much is spent on such policies (see the mean of the dep. variable row). When it comes to spending on assistance to Blacks, we see very little difference across education groups. When it comes to spending on improvement of Blacks, we observe that university graduates are significantly less likely to oppose such policy (the coefficient in column (6) is less than half the coefficient in column (5)). Yet, there is little difference between high school graduates and those who did not finish High School. As such, Table D.10 provides moderate evidence in favour of a policy impact, but suggests that information could still explain half of white men's anger.

Table D.10: Policy attitudes towards assistance to Blacks by levels of education

	(1)	(2) Too much on assistance	(3)	(4)	(5) Too much for improvement	(6)
White Male	0.033^{*}	0.029^{***}	0.028^{***}	0.068^{***}	0.074^{***}	0.035^{***}
	(0.058)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Sample	No qualif./answer	High School	University	No qualif./answer	High School	University
Mean dep variable	0.07	0.07	0.05	0.10	0.09	0.07
Individual controls	✓	✓	✓	✓	✓	✓
N.obs	1605	7549	5442	1605	7549	5442

Notes: Dependent variables are indicator variables, which equal one if respondent states that the US spends too much on improving the conditions of Blacks (columns (1)-(3)) or on assistance to Blacks (columns (4)-(6)). Complete model results can be found in Table F.11 in the Angry White Males - Dataverse.pdf file available in the APSR Dataverse for this article. Robust standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

When I look at the same survey items by age groups in Table D.11, we no longer see patterns consistent with a policy effect. For spending on the assistance to Blacks, there are very little differences between age groups (see columns (1)-(3)). For spending on the improvement of Blacks, over-65 white men differ more than any other age groups (columns (4)-(6)). This suggests again that information is a more likely cause of white men's anger, at least in the survey data analyzed in this appendix.

Table D.11: Policy attitudes towards assistance to Blacks by age groups

	(1)	(2) Too much on assistance	(3)	(4)	(5) Too much for improvement	(6)
White Male	0.027^{**}	0.031^{***}	0.028^{***}	0.012	0.056^{***}	0.076^{***}
	(0.034)	(0.000)	(0.004)	(0.337)	(0.000)	(0.000)
Sample	Under 25	26-64	Over 65	Under 25	26-64	Over 65
Mean dep variable	0.03	0.06	0.07	0.04	0.08	0.10
Individual controls	✓	✓	✓	✓	✓	✓
N.obs	1260	9824	3512	1260	9824	3512

Notes: Dependent variables are indicator variables, which equal one if respondent states that the US spends too much on improving the conditions of Blacks (columns (1)-(3)) or on assistance to Blacks (columns (4)-(6)). Complete model results can be found in Table F.12 in the Angry White Males - Dataverse.pdf file available in the APSR Dataverse for this article. Robust standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Cooperative Election Study

Using the Copperative Election Study, I look in Table D.12 how white men differ from other respondents on four items: feeling about white advantage or lack thereof (columns (1) and (2)), belief that racial problems are rare (columns (3) and (4)), belief that Blacks should work their way up without help (labelled resentment 1 in columns (5) and (6)) or that slavery and discrimination are not impeding Blacks' advancement (labelled resentment 2 in columns (7) and (8)). On all survey items, with or without controls, white men are more opposed to social changes than other respondents.

In Table D.13, I look at the first two items from Table D.12 (no advantages for Whites and no racial problems) by levels of education. Education reduces the willingness to say that whites have no advantage or that racial problems are rare as for other surveys (see the row mean of dep. variable). Yet, the coefficient on white men remains constant when we look at high school graduates and university graduates (the coefficient on no high school is actually lower when it comes to racial problems). Hence, there is little evidence in favour of a policy effect and, rather, some evidence in favour of information being the cause of white men's anger.

Table D.12: Policy attitudes towards racial discrimination

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	No advantages for Whites		Racial problems uncommon		Racial resentment 1		Racial resentment 2	
White Male	0.119***	0.103***	0.110***	0.104***	0.124***	0.101***	0.121***	0.091***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sample	All	All	All	All	All	All	All	All
Mean dep variable	0.29	0.29	0.22	0.22	0.51	0.51	0.40	0.40
Individual controls		✓		✓		✓		✓
N.obs	206864	206319	203284	202762	202873	202167	202888	202183

Notes: Dependent variables are indicator variables. For columns (1) and (2), variable equals one if respondent disagrees (strongly or somewhat) that Whites have advantages. For columns (3) and (4), variable equals one if respondent agrees (somewhat or strongly) that racial problems are rare. For columns (5) and (6), variable equals one if respondent agrees (somewhat or strongly) that Blacks should overcome prejudice without special favors. For columns (7) and (8), variable equals one if respondent disagrees (somewhat or strongly) that slavery and discrimination have created conditions that make it difficult for Blacks to progress socially. Complete model results can be found in Table F.13 in the Angry White Males - Dataverse.pdf file available in the APSR Dataverse for this article. Robust standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.13: Policy attitudes towards racial discrimination by levels of education

	(1)	(2)	(3)	(4)	(5)	(6)
Equal opport. to		No advantages for Whites			Racial problems uncommon	
White Male	0.111***	0.106***	0.101***	0.070***	0.105***	0.106***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sample	No high school	High School	University	No high school	High School	University
Mean dep variable	0.35	0.34	0.25	0.22	0.23	0.21
Individual controls	✓	✓	✓	✓	✓	✓
N.obs	5210	96490	104619	5086	94404	103272

Notes: Dependent variables are indicator variables. For columns (1) to (3), variable equals one if respondent disagrees (strongly or somewhat) that Whites have advantages. For columns (4) to (6), variable equals one if respondent agrees (somewhat or strongly) that racial problems are rare. Complete model results can be found in Table F.14 in the Angry White Males - Dataverse.pdf file available in the APSR Dataverse for this article. Robust standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

When we look by age groups in Table D.14, we see some evidence in favour of a policy effect. Over 65 white men look more similar to other respondents than other age groups. Yet, the coefficient on White Male for over 65 is only around 25% smaller than the coefficient for other age groups. Hence, while there is some evidence in favour of a policy effect, there is still some room for an informational source of white men's anger.

Table D.14: Policy attitudes towards racial discrimination by age groups

	(1)	(2)	(3)	(4)		
Equal opport. to		No advantages for Whites			Racial problems uncommon	
White Male	0.111***	0.112***	0.083***	0.111***	0.112***	0.086***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sample	Under 25	26-64	Over 65	Under 25	26-64	Over 65
Mean dep variable	0.15	0.29	0.34	0.16	0.22	0.22
Individual controls	✓	✓	✓	✓	✓	✓
N.obs	12944	143263	50112	12875	141305	48582

Notes: Dependent variables are indicator variables. For columns (1) to (3), variable equals one if respondent disagrees (strongly or somewhat) that Whites have advantages. For columns (4) to (6), variable equals one if respondent agrees (somewhat or strongly) that racial problems are rare. Complete model results can be found in Table F.15 in the Angry White Males - Dataverse.pdf file available in the APSR Dataverse for this article. Robust standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.