Online Appendix to A Formal Theory of Public Opinion

Daniel Diermeier* and Michael Schnabel†

April 21, 2024

A. Proofs of Main Propositions

Proposition 1 *The stationary distribution of the process of opinion formation is*

$$P_i(s_i) = \frac{e^{\beta a_i s_i}}{1 + e^{\beta a_i}}$$

with mean

$$\mu_i = \frac{1}{1 + e^{-\beta a_i}} = r_{i,\uparrow}$$

and variance

$$\sigma_i^2 = \mu_i - \mu_i^2 = \mu_i(1 - \mu_i)$$

where $0 \leq \mu_i \leq 1$.

Proof of Proposition 1: Let $P_i(1)$ and $P_i(0)$ denote the probabilities of voter i being in the state 1 or 0. These can be obtained from the condition

$$\begin{aligned} \begin{pmatrix} P_i(1) \\ P_i(0) \end{pmatrix} &= \begin{pmatrix} P_{i,\Delta t}(1|1) & P_{i,\Delta t}(1|0) \\ P_{i,\Delta t}(0|1) & P_{i,\Delta t}(0|0) \end{pmatrix} \begin{pmatrix} P_i(1) \\ P_i(0) \end{pmatrix} \\ &= \begin{pmatrix} 1 - r_{i,\downarrow}\Delta t & r_{i,\uparrow}\Delta t \\ r_{i,\downarrow}\Delta t & 1 - r_{i,\uparrow}\Delta t \end{pmatrix} \begin{pmatrix} P_i(1) \\ P_i(0) \end{pmatrix} \end{aligned}$$

which is a balance equation that any stationary probability distribution necessarily has to satisfy regardless of the particular value of Δt. The solution of this eigenvalue problem is

* Department of Political Science and Owen School of Management, Vanderbilt University, Email: daniel.diermeier@vanderbilt.edu.

† Department of Political Science, Vanderbilt University, Email: michael.schnabel@vanderbilt.edu.

provided by

$$\begin{pmatrix} P_i(1) \\ P_i(0) \end{pmatrix} = \begin{pmatrix} r_{i,\uparrow} \\ r_{i,\downarrow} \end{pmatrix},$$

which does not depend on Δt. Hence, we can write $P_i(s_i)$ as follows

$$P_i(s_i) = \frac{e^{\beta a_i s_i}}{1 + e^{\beta a_i}}$$

where s_i can assume the values 1 and 0. That is,

$$P_i(s_i = 0) = \frac{1}{1 + e^{\beta a_i}}, \qquad P_i(s_i = 1) = \frac{e^{\beta a_i}}{1 + e^{\beta a_i}} = \frac{1}{1 + e^{-\beta a_i}}.$$

We can then calculate the mean and variance of s_i with respect to the stationary distribution above and obtain for the mean

$$\mu_i = 0 \cdot P_i(s_i = 0) + 1 \cdot P_i(s_i = 1) = \frac{1}{1 + e^{-\beta a_i}} = r_{i,\uparrow},$$

and for the variance

$$\sigma_i^2 = E[(s_i - \mu)^2] = E[s_i^2] - \mu_i^2$$

with

$$E[s_i^2] = P_i(s_i = 0) \cdot 0 + P_i(s_i = 1) \cdot 1 = E[s_i] = \mu_i$$

and hence

$$\sigma_i^2 = \mu_i - \mu_i^2 = \mu_i(1 - \mu_i).$$

Note that $0 \leq \mu_i \leq 1$ follows from $\mu = r_{i,\uparrow}$ and the assumption that $0 \leq r_{i,\uparrow} \leq 1$. **QED**

Proposition 2 *For all $B \neq -a_i$: $\partial \chi_i(\beta, a_i, B)/\partial \beta$ is non-monotonic, first increasing in β until a maximum is reached at $\beta_*(a_i, B)$, and subsequently decreasing in β.*

Proof of Proposition 2: For any given a_i and B, $B \neq -a_i$ we define

$$\tilde{\beta} := \beta \cdot |a_i + B|, \quad \tilde{\beta} > 0$$

which allows us to write β as

$$\beta = \tilde{\beta}/|a_i + B|$$

such that

$$\chi_i(\beta, a_i, B) = \frac{\beta\, e^{-\beta(a_i+B)}}{\left(1+e^{-\beta(a_i+B)}\right)^2} = \frac{\tilde{\beta}}{|a_i+B|}\frac{e^{-\tilde{\beta}\epsilon}}{\left(1+e^{-\tilde{\beta}\epsilon}\right)^2}$$

where $\epsilon := \operatorname{sign}(a_i + B) \in \{-1, 1\}$.
The last expression gives the same value for both, $\epsilon = +1$ or $\epsilon = -1$ and therefore

$$\chi_i(\beta, a_i, B) = \frac{1}{|a_i+B|}\frac{\tilde{\beta}\, e^{-\tilde{\beta}}}{\left(1+e^{-\tilde{\beta}}\right)^2}.$$

Because the change from β to $\tilde{\beta}$ is merely a rescaling of β, it suffices to show that the last expression has a unique maximum at some $\tilde{\beta}_*$. This depends on the properties of the function

$$f(\tilde{\beta}) = \frac{\tilde{\beta}\, e^{-\tilde{\beta}}}{\left(1+e^{-\tilde{\beta}}\right)^2}.$$

Note that $f(\tilde{\beta})$ can be written as the product

$$f(\tilde{\beta}) = \tilde{\beta} \cdot g(\tilde{\beta})$$

where $g(\tilde{\beta})$ denotes the *logistic distribution* $g(\tilde{\beta}) = \frac{e^{-\tilde{\beta}}}{\left(1+e^{-\tilde{\beta}}\right)^2} \geq 0$. At $\tilde{\beta} = 0$ the function $f(\tilde{\beta})$ starts out at zero, $f(\tilde{\beta} = 0) = 0$, and has a positive slope $f'(0) = 1/4$ which follows from

$$f'(\tilde{\beta}) = \frac{df(\tilde{\beta})}{d\tilde{\beta}} = g(\tilde{\beta}) + \tilde{\beta} \cdot \frac{dg(\tilde{\beta})}{d\tilde{\beta}}$$

and $g(0) = 1/4$. Maxima of $f(\tilde{\beta})$ have vanishing slope $f'(\tilde{\beta}) = 0$, and occur whenever the condition

$$-\frac{dg(\tilde{\beta}_*)}{d\tilde{\beta}_*}\frac{1}{g(\tilde{\beta}_*)} = \frac{1}{\tilde{\beta}_*}$$

is met, which can be recast to

$$\tanh\left(\tilde{\beta}_*/2\right) = 1/\tilde{\beta}_*$$

Geometrically these correspond to the values of $\tilde{\beta} > 0$ at which the sigmoid function $\tanh\left(\tilde{\beta}/2\right)$ and the hyperbola $1/\tilde{\beta}$ intersect, which happens *exactly once*, because $\tanh\left(\tilde{\beta}/2\right)$ is monotonously increasing (approaching 1 for $\tilde{\beta} \to \infty$) while $1/\tilde{\beta}$ is monotonously decreasing (approaching 0 for $\tilde{\beta} \to \infty$). The solution can be found numerically and reads $\tilde{\beta}_* \approx 1.5434$.

The maximum susceptibility then reads

$$\chi_i^*(\beta_*) = \frac{\tilde{\beta}_*}{|a_i + B|} \frac{e^{-\tilde{\beta}_*}}{\left(1 + e^{-\tilde{\beta}_*}\right)^2} = \beta_* \cdot \gamma$$

where the constant γ is defined by

$$\gamma = \frac{e^{-\tilde{\beta}_*}}{\left(1 + e^{-\tilde{\beta}_*}\right)^2} \approx 0.14505$$

and corresponds to the slope of *dashed line* in Figure 2. The fact that all maxima are located on a line $\chi_i^*(\beta_*) \sim c \cdot \beta_*$ is a consequence of the fact that the susceptibility is a function of $\beta(a_i + B)$ and will also hold for a broad class of sigmoidal update functions. For more details we refer the reader to Diermeier and Schnabel (2024).

QED

Corollary 2.1 *We can also show for a fixed β a voter's susceptibility will be maximal for $B = -a_i$ with*

$$\chi_i(\beta, a_i, B = -a_i) = \beta/4.$$

Thus $\beta/4$ constitutes an upper bound for the susceptibility, for any a_i and B.

Proof of Corollary 2.1: This follows from writing the susceptibility as

$$\chi_i(\beta, a_i, B) = \beta \cdot g(\beta(a_i + B))$$

where $g(x)$ denotes the logistic distribution

$$g(x) = \frac{e^{-x}}{(1 + e^{-x})^2}$$

which has a global maximum at $x = 0$, i.e.

$$g(x) < g(0) = 1/4, \forall x \neq 0.$$

Hence

$$\chi_i(\beta, a_i, B) < \chi_i(\beta, a_i, B = -a_i) = \beta/4, \forall B \neq -a_i.$$

QED

Proposition 3 *Suppose that individuals form their opinions independently from each other and that there are two fractions of the population such that for all a_i either $a_i = a_+ > 0$, or $a_i = a_- < 0$. Let f_+ denote the fraction of the population with $a_i = a_+$ and f_- the fraction of the population with $a_i = a_-$. Suppose the transition rates are such that for all i with $a_i = a_+$,*

$$r_{+,\uparrow} = \frac{1}{1+e^{-\beta a_+}} =: \mu_+$$

and

$$r_{+,\downarrow} = \frac{1}{1+e^{\beta a_+}},$$

and for all i with $a_i = a_-$

$$r_{-,\uparrow} = \frac{1}{1+e^{-\beta a_-}} =: \mu_-$$

and

$$r_{-,\downarrow} = \frac{1}{1+e^{\beta a_-}}.$$

Then the aggregate opinion x is distributed as a normal distribution,

$$P_N(x) = \frac{1}{\sqrt{2\pi\sigma^2(N)}} e^{-(x-\mu(N))^2/(2\sigma^2(N))}$$

with mean

$$\mu(N) = f_- \cdot \mu_- + f_+ \cdot \mu_+$$

and variance

$$\sigma^2(N) = \frac{1}{N}\left(f_- \cdot \sigma_-^2 + f_+ \cdot \sigma_+^2\right).$$

Proof of Proposition 3: The central limit theorem states that the average $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ of a sufficiently large number of independent random numbers y_i with *finite* variance $\sigma_{y_i}^2 < \infty$ will approach a normal distribution in the large N limit, $N \to \infty$. In our case with s_i as independent random variables and

$$P_i(s_i) = \frac{1}{1+e^{-\beta a_i(2s_i-1)}}$$

we know that

$$\mu_i = E[s_i] = \frac{1}{1+e^{-\beta a_i}}$$

$$\sigma_i^2 = \mu_i(1-\mu_i) = \frac{1}{(1+e^{-\beta a_i})(1+e^{\beta a_i})} \leq 1/4.$$

Hence, the condition of the central limit theorem is satisfied and, for $N \to \infty$, the aggregate opinion

$$x = \frac{1}{N}\sum_{i=1}^{N} s_i \tag{1}$$

is normally distributed and described by a steady state distribution $P_N(x)$. Moreover, this will also be the case for

$$x(t) = \frac{1}{N}\sum_{i=1}^{N} s_i(t) \tag{2}$$

at any time t. Assuming the parameters β and a_i are constant in time, the corresponding distribution is given by

$$P_N(x) = \frac{1}{\sqrt{2\pi\sigma^2(N)}} e^{-(x-\mu(N))^2/(2\sigma^2(N))}$$

where we still need to calculate the mean $\mu(N) = E[x]$ and variance $\sigma^2(N) = E[x^2] - \mu^2(N)$. By means of the definitions Eq.(1) or Eq.(2), $\mu(N)$ and $\sigma^2(N)$ can be expressed in terms of s_i as follows

$$\mu(N) = E[x] = \frac{1}{N}\sum_{i=1}^{N} E[s_i(t)] = \frac{1}{N}\sum_{i=1}^{N} s_i P_i(s_i) = \frac{1}{N}\sum_{i=1}^{N} P_i(s_i = 1) = \frac{1}{N}\sum_{i=1}^{N} \mu_i$$

where

$$\mu_i = r_{i,\uparrow}$$

denotes the average opinion of an individual i. In the case of polarized populations we have

$$\mu(N) = E[x] = \frac{1}{N}\sum_{i=1}^{N} E[s_i(t)] = \frac{1}{N}\left(\sum_{i=1}^{N}\sum_{x_i=0}^{1} s_i f_- P_i(s_i; a_-) + s_i f_+ P_i(s_i; a_+)\right).$$

Then, because

$$\mu_- = \sum_{s_i=0}^{1} s_i P_i(s_i; a_-)$$

and

$$\mu_+ = \sum_{s_i=0}^{1} s_i P_i(s_i; a_+)$$

we get

$$\begin{aligned}\mu(N) &= \frac{1}{N}\sum_{i=1}^{N}(f_-\mu_- + f_+\mu_+)\\ &= f_-\mu_- + f_+\mu_+.\end{aligned}$$

The population variance for heterogeneous a_i is defined by

$$\sigma^2(N) = E[x^2] - \mu^2(N)$$

and calculated next. The quantity x^2 can be expressed in terms of s_i as follows

$$x^2 = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} s_i(t)s_j(t).$$

Taking the expectation on both sides,

$$\begin{aligned}E[x^2] &= \frac{1}{N^2}E[\sum_{i=1}^{N}\sum_{j=1}^{N} s_i(t)s_j(t)] = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} E[s_i(t)s_j(t)]\\ &= \frac{1}{N^2}\sum_{i=1}^{N} E[s_i(t)s_i(t)] + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} E[s_i(t)s_j(t)]\end{aligned}$$

and using

$$E[s_i(t)s_j(t)] = E[s_is_j] = E[s_i]E[s_j]$$

for $i \neq j$ (due to the assumed time-independence of a_i and β and the statistical independence of x_i and x_j) and

$$E[s_i(t)s_i(t)] = E[s_is_i] = E[s_i]$$

(because $s_i^2 = s_i$ regardless of whether $s_i = 0$ or $s_i = 1$) we obtain

$$\begin{aligned}E[x^2] &= \frac{1}{N^2}\sum_{i=1}^{N} E[s_i] + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} E[s_i]E[s_j]\\ &= \frac{1}{N^2}\sum_{i=1}^{N} E[s_i] + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} E[s_i]E[s_j] - \frac{1}{N^2}\sum_{i=1}^{N} E[s_i]^2.\end{aligned}$$

Because

$$\frac{1}{N}\sum_{i=1}^{N} E[s_i] = \mu(N)$$

we get

$$E[x^2] = \frac{\mu(N)}{N} + \mu^2(N) - \frac{1}{N^2}\sum_{i=1}^{N} E[s_i]^2$$

such that

$$\sigma^2(N) = E[x^2] - \mu^2(N) = \frac{\mu(N)}{N} - \frac{1}{N^2}\sum_{i=1}^{N} E[s_i]^2$$

or

$$\sigma^2(N) = \frac{\mu(N)}{N} - \frac{1}{N^2}\sum_{i=1}^{N} \mu_i^2$$

Then, because $a_i = a_+$ for a fraction f_+ and $a_i = a_-$ for fraction f_- in the population we obtain

$$\begin{aligned}\sigma^2(N) &= \frac{\mu(N)}{N} - \frac{1}{N^2}\sum_{i=1}^{N_-} \mu_-^2 - \frac{1}{N^2}\sum_{i=1}^{N_+} \mu_+^2 = \frac{f_-\mu_-}{N} - \frac{Nf_-}{N^2}\mu_-^2 + \frac{f_+\mu_+}{N} - \frac{Nf_+}{N^2}\mu_+^2 \\ &= \frac{f_-\mu_-(1-\mu_-)}{N} + \frac{f_+\mu_+(1-\mu_+)}{N} \\ &= \frac{1}{N}\left(f_-\sigma_-^2 + f_+\sigma_+^2\right).\end{aligned}$$

QED

Corollary 3.1 *In the case where both* $\sigma_-^2 > 0$ *and* $\sigma_+^2 > 0$ *there exists an* N' *such that for all* $N > N' : \sigma^2(N) < \min\{\sigma_+^2, \sigma_-^2\}$.

Proof of Corollary 3.1: Proposition 3 and the assumption that $\sigma_\pm^2 > 0$ implies that $\sigma^2(N) = \frac{1}{N}\left(f_-\sigma_-^2 + f_+\sigma_+^2\right)$ where $f_- = (1 - f_+)$ such that

$$0 < C_{min} =: \min\{\sigma_+^2, \sigma_-^2\} \leq f_-\sigma_-^2 + f_+\sigma_+^2 \leq \max\{\sigma_+^2, \sigma_-^2\} := C_{max}$$

Furthermore, from $\sigma_i^2 = \mu_i(1-\mu_i)$ we know that $C_{max} \leq 1/4$ and therefore finite. It follows that

$$\sigma^2(N') = \frac{f_-\sigma_-^2 + f_+\sigma_+^2}{N'} \leq \frac{C_{max}}{N'}$$

and therefore

$$\sigma^2(N') \leq C_{min}$$

for

$$\frac{C_{max}}{N'} \leq C_{min}$$

which is fulfilled for

$$N' \geq \frac{C_{max}}{C_{min}}.$$

QED

Proposition 4 *The polarization index q for a bipartite population with negative and positive attitudes varies within*

$$0 \leq q \leq 1.$$

Proof of Proposition 4: Substituting $f_- = 1 - f_+$ and $\mu = f_+\mu_+ + (1 - f_+)\mu_-$ in the definition of q, one obtains

$$q = 4(1 - f_+)f_+(\mu_+ - \mu_-).$$

Because

$$0 \leq \mu_- \leq \mu_+ \leq 1$$

the bracket $(\mu_+ - \mu_-)$ varies within $0 \leq (\mu_+ - \mu_-) \leq 1$. Furthermore, the expression $4(1 - f_+)f_+$ examined as a polynomial in f_+ is 0 for $f_+ = 0$ and $f_+ = 1$ and maximal at $f_+ = 1/2$ where it becomes 1.

QED

Proposition 5 *For all $a_+ > 0, a_- < 0$ the polarization index $q(a_+, a_-; \beta)$ in a polarized population increases with β,*

$$\frac{\partial q(a_+, a_-; \beta)}{\partial \beta} > 0$$

and for $\beta \to \infty$ we have $q(a_+, a_-; \beta) \to 4(1 - f_+)f_+$.

Proof of Proposition 5: From the definition of q we have

$$q(a_+, a_-; \beta) = 4(1 - f_+)f_+ \cdot (\frac{1}{1 + e^{-\beta a_+}} - \frac{1}{1 + e^{-\beta a_-}}) = 4(1 - f_+)f_+ \cdot (\mu_+ - \mu_-)$$

where $\mu_+ > \mu_-$ because $a_+ > a_-$. Taking the derivative with respect to β one obtains

$$\begin{aligned}\frac{\partial q(a_+, a_-; \beta)}{\partial \beta} &= 4(1 - f_+)f_+ \cdot \left(\frac{a_+ e^{-\beta a_+}}{(1 + e^{-\beta a_+})^2} - \frac{a_- e^{-\beta a_-}}{(1 + e^{-\beta a_-})^2}\right) \\ &= 4(1 - f_+)f_+ \cdot (a_+ \mu_+^2 e^{-\beta a_+} - a_- \mu_-^2 e^{-\beta a_-}).\end{aligned}$$

Therefore, $\partial q(a_+, a_-; \beta)/\partial\beta > 0$ because $a_- < 0$, and $-a_- > 0$ as assumed. Furthermore, because

$$\lim_{\beta\to\infty} \frac{1}{1+e^{-\beta a_+}} = 1$$

and

$$\lim_{\beta\to\infty} \frac{1}{1+e^{-\beta a_-}} = 0$$

one finds that

$$\lim_{\beta\to\infty} q(a_+, a_-; \beta) = 4(1-f_+)f_+ = 4f_-f_+.$$

Corollary 5.1 *For the special case where* $f_+ = 1/2$ *it follows that*

$$\lim_{\beta\to\infty} q(a_+, a_-; \beta) \to 1.$$

Proof of Corollary 5.1: This follows directly from Proposition 5.

QED

Corollary 5.2 *For all* $a_+ > 0, a_- < 0 : q(a_+, a_-; \beta)$

$$\frac{\partial^2 q(a_+, a_-; \beta)}{\partial^2\beta} < 0.$$

Proof of Corollary 5.2: From

$$\frac{\partial q(a_+, a_-; \beta)}{\partial\beta} = 4(1-f_+)f_+ \cdot (a_+\mu_+^2 e^{-\beta a_+} - a_-\mu_-^2 e^{-\beta a_-})$$

it follows that

$$\frac{\partial^2 q(a_+, a_-; \beta)}{\partial^2\beta} = -4(1-f_+)f_+ \left(a_+^2\mu_+^3 e^{-\beta a_+}(1-e^{-\beta a_+}) - a_-^2\mu_-^3 e^{-\beta a_-}(1-e^{-\beta a_-})\right)$$

which will be negative because $(1-e^{-\beta a_+}) > 0$ and $(1-e^{-\beta a_-}) < 0$ due to our assumption that $a_+ > 0$ and $a_- < 0$.

QED

B. Supplementary Propositions

Proposition S1 *An increase in the magnitude of positive considerations shifts the mean of the opinion distribution in a positive direction towards expressing 1, while increasing the magnitude of negative considerations shifts the mean in a negative direction towards expression 0. Formally,*

$$\frac{\partial \mu_i}{\partial |c_k^{i,+}|} > 0, \text{ and } \frac{\partial \mu_i}{\partial |c_k^{i,-}|} < 0.$$

Proof of Proposition S1: Note that

$$a_i = \sum_K w_k^i c_k^i = \sum_{K^+} w_k^{i,+} c_k^{i,+} + \sum_{K^-} w_k^{i,-} c_k^{i,-} = \sum_{K^+} w_k^{i,+} |c_k^{i,+}| - \sum_{K^-} w_k^{i,+} |c_k^{i,-}|.$$

Now rewrite μ_i as follows

$$\mu_i = \frac{1}{1 + e^{-\beta a_i}} = \frac{1}{1 + \exp[-\beta(\sum_{K^+} w_k^{i,+} |c_k^{i,+}| - \sum_{K^-} w_k^{i,-} |c_k^{i,-}|)]}$$

From this follows that

$$\frac{\partial \mu_i}{\partial |c_k^{i,+}|} = \frac{\partial \mu_i}{\partial a_i} \frac{\partial a_i}{\partial |c_k^{i,+}|} = e^{-\beta a_i} \beta \mu_i^2 w_k^{i,+} > 0, \text{ and } \frac{\partial \mu_i}{\partial |c_k^{i,-}|} = \frac{\partial \mu_i}{\partial a_i} \frac{\partial a_i}{\partial |c_k^{i,-}|} = -e^{-\beta a_i} \beta \mu_i^2 w_k^{i,-} < 0,$$

where we used that

$$\frac{\partial \mu_i}{\partial a_i} = \frac{\partial}{\partial a_i} \left(\frac{1}{1 + e^{-\beta a_i}} \right) = \frac{\beta e^{-\beta a_i}}{(1 + e^{-\beta a_i})^2} = \beta e^{-\beta a_i} \mu_i^2.$$

QED

Corollary S1.1 *If positive considerations become more accessible, mean opinions will shift in a positive direction; if negative considerations become more accessible, the expression of negative opinions will become more likely. Formally,*

$$\frac{\partial \mu_i}{\partial w_k^{i,+}} > 0, \text{ and } \frac{\partial \mu_i}{\partial w_k^{i,-}} < 0.$$

Proof of Corollary S1.1: The result follows directly from

$$\frac{\partial \mu_i}{\partial w_k^{i,+}} = \frac{\partial \mu_i}{\partial a_i}\frac{\partial a_i}{\partial w_k^{i,+}} = e^{-\beta a_i}\mu_i^2 |c_k^{i,+}| > 0, \text{ and } \frac{\partial \mu_i}{\partial w_k^{i,-}} = \frac{\partial \mu_i}{\partial a_i}\frac{\partial a_i}{\partial w_k^{i,-}} = -e^{-\beta a_i}\mu_i^2 |c_k^{i,-}| < 0.$$

QED

Corollary S1.2 *Increased salience increases the strength of mean opinions. Formally,*

$$\frac{\partial \mu_i}{\partial \beta} = \begin{cases} > 0 & \text{if } a_i > 0 \\ < 0 & \text{if } a_i < 0 \end{cases}.$$

Proof of Corollary S1.2: The result follows directly from

$$\frac{\partial \mu_i}{\partial \beta} = \frac{\partial}{\partial \beta}\left(\frac{1}{1+e^{-\beta a_i}}\right) = \mu_i^2 a_i\, e^{\beta a_i}.$$

QED

Corollary S1.3 *Increasing the magnitude or salience of positive considerations will decrease variance provided that the overall attitude is positive. Otherwise, variance will increase. Formally,*

$$\frac{\partial \sigma_i^2}{\partial |c_k^{i,+}|} < 0 \text{ and } \frac{\partial \sigma_i^2}{\partial w_k^{i,+}} < 0 \text{ if } a_i = \sum_{K^+} w_k^{i,+}|c_k^{i,+}| - \sum_{K^-} w_k^{i,-}|c_k^{i,-}| > 0$$

and

$$\frac{\partial \sigma_i^2}{\partial |c_k^{i,+}|} > 0 \text{ and } \frac{\partial \sigma_i^2}{\partial w_k^{i,+}} > 0 \text{ if } a_i = \sum_{K^+} w_k^{i,+}|c_k^{i,+}| - \sum_{K^-} w_k^{i,-}|c_k^{i,-}| < 0.$$

Proof of Corollary S1.3: The result follows directly from $\sigma_i^2 = \mu_i(1-\mu_i)$ and

$$\frac{\partial \sigma_i^2}{\partial |c_k^{i,+}|} = \frac{\partial \sigma_i^2}{\partial a_i}\frac{\partial a_i}{\partial |c_k^{i,+}|}$$

where $a_i = \sum_{K^+} w_k^{i,+}|c_k^{i,+}| - \sum_{K^-} w_k^{i,-}|c_k^{i,-}|$. From the definition of a_i follows that $\partial a_i/\partial |c_k^{i,+}| = w_k^{i,+}$, which for an active consideration will be positive because $w_k^{i,+} > 0$. Hence it remains to evaluate the sign of $\partial \sigma_i^2/\partial a_i$. From

$$\sigma_i^2 = \mu_i(1-\mu_i) = \frac{1}{(1+e^{-\beta a_i})\,(1+e^{\beta a_i})}$$

follows

$$\frac{\partial \sigma_i^2}{\partial a_i} = \frac{a_i \left(1 - e^{\beta a_i}\right) e^{\beta a_i}}{\left(1 + e^{\beta a_i}\right)^3}$$

and $\partial \sigma_i^2 / \partial a_i < 0$ because $e^{\beta a_i} > 1$ for $\beta > 0$ and $a_i > 0$. Thus, $\partial \sigma_i^2 / \partial |c_k^{i,+}| < 0$. Likewise, with $\partial a_i / \partial w_k^{i,+} = |c_k^{i,+}|$ it follows that $\partial \sigma_i^2 / \partial w_k^{i,+} < 0$ if $a_i > 0$. The opposite will be the case for $a_i < 0$.

QED

Corollary S1.4 *As issue attention increases, variance will decrease for positive and negative attitudes if $\beta > 0$. Formally,*

$$\frac{\partial \sigma^2}{\partial \beta} < 0 \text{ if } a_i \neq 0.$$

Proof of Corollary S1.4: This follows from the definition of $\sigma_i^2 = \mu_i(1 - \mu_i)$ and therefore

$$\frac{\partial \sigma^2}{\partial \beta} = \frac{\partial \mu_i}{\partial \beta}(1 - 2\mu_i).$$

When $\beta a_i > 0$ we have $\mu_i > 1/2$, $\partial \mu_i / \partial \beta > 0$ and therefore $\partial \sigma^2 / \partial \beta > 0$. Likewise, for $\beta a_i < 0$ we have $\mu_i < 1/2$, $\partial \mu_i / \partial \beta < 0$ and again $\partial \sigma^2 / \partial \beta > 0$. **QED**

Corollary S1.5 *("Uncertainty"): Assuming that a single consideration, c_+^i, is activated ($K = 1$), here assumed to have a positive valence, $c_+^i > 0$. When engagement increases uncertainty diminishes and average opinion strengthens. Formally, when $K = 1$ and $c_+^i > 0$,*

$$\frac{\partial \sigma_i^2}{\partial \beta} < 0 \quad \text{and} \quad \frac{\partial \mu_i}{\partial \beta} > 0$$

Proof of Corollary S1.5: The result follows directly from $\sigma_i^2 = \mu_i(1 - \mu_i)$ so that

$$\frac{\partial \sigma_i^2}{\partial \beta} = \frac{\partial}{\partial \beta}(\mu_i(1 - \mu_i)) = \frac{\partial \mu_i}{\partial \beta}(1 - 2\mu_i)$$

Since $K = 1$ we have $w^{i,+} = 1$ and therefore $a_i = c^{i,+} > 0$. From the definition of μ_i and Corollary S1.2 then follows that $\mu_i > 1/2$ as well as $\partial \mu_i / \partial \beta > 0$. This implies that $\partial \sigma_i^2 / \partial \beta < 0$. **QED**

Corollary S1.6 *("Equivocation"): Adding a second consideration, $c^{i,++}$, whose valence has the same sign as the first, i.e., $c^{i,++} > 0$, uncertainty diminishes and average opinion strengthens when engagement in increases. If $|c^{i,++}| > |c^{i,+}|$ then the shifts in mean opinion and the reductions in response variability as a function of β will be more pronounced.*

Proof of Corollary S1.6: With two activated considerations, $c^{i,+} > 0$, $c^{i,++} > 0$ and $w^{i,+} > 0$, $w^{i,++} > 0$ the attitude becomes $\tilde{a}_i = w^{i,+}c^{i,+} + w^{i,++}c^{i,++} > 0$ and is positive. Therefore, $\partial\tilde{\mu}_i/\partial\beta > 0$ and $\partial\tilde{\sigma}_i^2/\partial\beta < 0$ just as shown in Corollary 4b. For $|c^{i,++}| > |c^{i,+}|$ we find that $\tilde{a}_i > a_i$ since

$$\tilde{a}_i = w^{i,+}c^{i,+} + w^{i,++}c^{i,++} = w^{i,+}c^{i,+} + (1 - w^{i,+})c^{i,++} > c^{i,+} = a_i$$

due to the normalization $w^{i,+} + w^{i,++} = 1$. From $\tilde{a}_i > a_i$ then follows that $\tilde{\mu}_i \geq \mu_i$ and therefore also $\tilde{\sigma}_i^2 \leq \sigma_i^2$ since $\mu_i(a_i) = \frac{1}{1+e^{-\beta a_i}}$ is monotonic in a_i, and due to the concavity of $\sigma_i^2(a_i) = \mu_i(a_i) \cdot (1 - \mu_i(a_i))$ which has its global maximum at $a_i = 0$ where $\sigma_i^2(0) = 1/4$ **QED**

Corollary S1.7 *("Ambivalence"): Adding a second consideration with an opposite valence, $c^{i,-} < 0$, variance increases and average opinion weakens, as long as the second opinion doesn't override the first one, i.e. $w^{i,-}|c^{i,-}| < 2w^{i,+}|c^{i,+}|$. Formally, if $w^{i,-}|c^{i,-}| < 2w^{i,+}|c^{i,+}|$ we have $\mu(\tilde{a}_i) \leq \mu(a_i)$ and $\sigma^2(\tilde{a}_i) \geq \sigma^2(a_i)$ where $\tilde{a}_i = w^{i,+}c^{i,+} + w^{i,-}c^{i,-}$ and $a_i = c^{i,+}$.*

Proof of Corollary S1.7: We compare the new attitude $\tilde{a}_i = w^{i,+}c^{i,+} + w^{i,-}c^{i,-}$ to the original one, $a_i = c^{i,+}$. For $w^{i,-}|c^{i,-}| < 2w^{i,+}|c^{i,+}|$ we have $|\tilde{a}_i| <|a_i|$, and therefore $\mu(\tilde{a}_i) \leq \mu(a_i)$ just as explained in the proof of the previous Corollary. This also implies $\sigma^2(|\tilde{a}_i|) \geq \sigma^2(|a_i|)$ due to the concavity of $\sigma^2(|a_i|)$. **QED**

Corollary S1.8 *The change of opinion variance upon an increase of β as described by $\partial\sigma_i^2/\partial\beta$ in our model reads*

$$\frac{\partial\sigma_i^2}{\partial\beta} = \frac{a_i \cdot e^{\beta a_i}\left(1 - e^{\beta a_i}\right)}{\left(1 + e^{\beta a_i}\right)^3}$$

Depending on the value of β and a_i this slope may be negative or zero, but not positive.

Proof of Corollary S1.8: By definition $\beta \geq 0$ and the denominator on the right side of the expression is always positive. For $\beta \cdot a_i = 0$ we find $\partial\sigma_i^2/\partial\beta = 0$ since $1 - e^{\beta a_i} = 0$ in the numerator. For $\beta \cdot a_i > 0$ the numerator is negative and therefore $\partial\sigma_i^2/\partial\beta < 0$. Likewise, $\beta a_i < 0$ implies negative $a_i < 0$ and positive$\left(1 - e^{\beta a_i}\right)$, such that the nominator is negative and therefore $\partial\sigma_i^2/\partial\beta < 0$. **QED**

Corollary S1.9 *In the extreme case of perfect ambivalence, i.e., $w^{i,+}c^{i,+} \approx w^{i,-}c^{i,-}$, we would have $a_i \approx 0$ and*

$$\frac{\partial\sigma_i^2}{\partial\beta} \approx 0.$$

Proof of Corollary S1.9: The case of perfect ambivalence is described by $w^{i,+}c^{i,+} \to -w^{i,-}c^{i,-}$ so that $a_i \to 0$. As explained in the previous Corollary $\partial\sigma_i^2/\partial\beta \to 0$ when $a_i \to 0$. A Taylor expansion of $\partial\sigma_i^2/\partial\beta$ around $a_i = 0$ yields

$$\partial\sigma_i^2/\partial\beta = -\frac{a_i^2\beta}{8} + \frac{a_i^4\beta^3}{24} + O\left(a_i^6\right)$$

so that $\partial\sigma_i^2/\partial\beta$ exhibits a quadratic maximum in a_i at $a_i = 0$. **QED**

Corollary S1.10 *The variance $\sigma_i^2(\beta, a_i)$, as a function of $\beta \geq 0$, is constant for $a_i = 0$*

$$\sigma_i^2(\beta, a_i = 0) = 1/4$$

and monotonic decreasing for $|a_i| > 0$ with a maximum at $\beta = 0$.

Proof of Corollary S1.10: This follows form the definition of $\sigma_i^2(\beta, a_i)$,

$$\sigma_i^2(\beta, a_i) = \mu_i(1 - \mu_i) = \frac{1}{\left(1 + e^{-\beta a_i}\right)\left(1 + e^{\beta a_i}\right)}$$

For $a_i = 0$ the numerator vanishes and we have $\sigma_i^2(\beta, a_i = 0) = 1/4$. For non-zero value of $\beta \cdot a_i$ the slope $\partial\sigma_i^2/\partial\beta$ is always negative as shown in Corollary S1.8. **QED**

Corollary S1.11 *Near $\beta = 0$, the larger $|a_i|$ the faster $\sigma^2(\beta, a_i)$ drops (as a function of β) which is captured by the negative curvature*

$$-\frac{\partial^2\sigma^2(\beta, a_i)}{\partial^2\beta}|_{\beta=0} = |a_i|^2/8$$

which increases as attitudes $|a_i|$ become more pronounced.

Proof of Corollary S1.11: This follows from the definition of $\partial\sigma_i^2/\partial\beta$ provided in Corollary S1.8 and taking the derivative with respect to β. **QED**

Proposition S2 *For populations in which the opposing attitudes are of comparable strengths, i.e. $a_+ = a = -a_-$ we have*

$$\mu_+ = r_{a,\uparrow} = \frac{1}{1+e^{-\beta a}} = r_{-a,\downarrow} = 1 - \mu_-$$

and

$$\mu_- = r_{-a,\uparrow} = \frac{1}{1+e^{\beta a}} = r_{a,\downarrow} = 1 - \mu_+$$

such that $\sigma_+^2 = \sigma_-^2$, because

$$\sigma_+^2 = \mu_+(1-\mu_+) = \mu_+\mu_- = \mu_-(1-\mu_-) = \sigma_-^2$$

and

$$\sigma_+^2 = \sigma_-^2 = r_{a,\uparrow} r_{a,\downarrow} = r_{a,\uparrow}(1 - r_{a,\downarrow}) = \frac{1}{1+e^{\beta a}} \frac{1}{1+e^{-\beta a}}$$

Henceforth, as long as $a_- = -a_+$ the population variance $\sigma^2(N)$ does *not* depend on f_+ and f_- and thus is invariant or robust with respect to the particular composition of the two subpopulations, and simply amounts to

$$\sigma^2(N) = \frac{1}{N}\left(f_-\sigma_-^2 + f_+\sigma_+^2\right) = \frac{\mu_+\mu_-}{N} = \frac{1}{N}\frac{1}{(1+e^{\beta a})(1+e^{-\beta a})}$$

As a consequence, the variance of expressed opinions is only sensitive to $\beta \cdot |a|$ regardless of the particular composition of the public. In particular, the expression above also applies to two special cases, (1) for balanced populations, that we define as referring to the situation where the relative group sizes are the same, i.e. $f_+ = f_- = 1/2$, and (2) for homogenous populations, in which either $f_+ = 1$ and $f_- = 0$, or vice versa. The results are summarized in the following corollaries.

Corollary S2.1 *For a balanced population with $f_+ = f_- = 1/2$ and $a_- = -a_+ = -a$ we have*

$$\mu(N) = 1/2$$

and

$$\sigma^2(N) = \frac{1}{N}\left(r_{a,\uparrow}\left(1 - r_{a,\uparrow}\right)\right) = \frac{1}{N}\frac{1}{(1+e^{\beta a})(1+e^{-\beta a})}.$$

Corollary S2.2 *For a homogenous population with $f_+ = 1$ and $f_- = 0$ and $a_+ = a$ one* *finds*

$$\mu(N) = \mu_+ = r_{a,\uparrow}$$

and

$$\sigma^2(N) = \frac{1}{N}\left(r_{a,\uparrow}\left(1 - r_{a,\uparrow}\right)\right) = \frac{1}{N} r_{a,\uparrow} r_{-a,\uparrow} = \sigma_i^2/N = \frac{1}{N}\frac{1}{(1+e^{\beta a})(1+e^{-\beta a})}.$$

Likewise, a homogenous population with $f_+ = 0$ *and* $f_- = 1$ *and* $a_- = -a$*,*

$$\mu(N) = \mu_- = r_{-a,\uparrow} = 1 - r_{a,\uparrow}$$

will have the same variance,

$$\sigma^2(N) = \frac{1}{N}\left(r_{-a,\uparrow}\left(1 - r_{-a,\uparrow}\right)\right) = \frac{1}{N} r_{-a,\uparrow} r_{a,\uparrow} = \sigma_i^2/N = \frac{1}{N}\frac{1}{(1+e^{\beta a})(1+e^{-\beta a})}.$$

The homogeneous case captures the situation where there is (nearly) *uniform consensus* on a policy, e.g., support for public funding of K-12 education. With $a_i = a$ for all i we simply write $r_{a,\uparrow}$ and $r_{a,\downarrow}$.

Note also that the population variance

$$\sigma^2(N) = \frac{\sigma_i^2}{N} = \frac{r_{a,\uparrow}\left(1 - r_{a,\uparrow}\right)}{N} = \frac{1}{N}\left(\frac{1}{1+e^{-\beta a}}\frac{1}{1+e^{\beta a}}\right)$$

is maximal for $\beta\, a = 0$, where it becomes $\sigma^2(N) = 1/(4N)$. This formally captures the "non-attitude" case. In contrast, $\sigma^2(N)$ vanishes as $\beta \cdot a \to \pm\infty$. Thus, our model predicts that the variance of public opinion decreases for strong attitudes (measured by high $|a|$) and higher engagement and attention (measured by β), and this relationship holds both for individual and aggregate opinion.[1]. But this is not the only effect. As in the individual case, the *mean* $\mu(N)$ of the distribution shifts in the direction of opinion 1 as $\beta \cdot a$ increases. As engagement and attention (measured by β) increase, aggregate opinion will shift towards the opinion favored by the sign of a.

As in the case of individual opinions, we can define attitude, attention, and susceptibility with similar properties. Thus susceptibility analysis in populations works similarly to the individual case, but is derived for each sub-population separately.

[1] For evidence at both the individual and aggregate level, see Alvarez and Brehm (2002), especially Chapter 8.

Proposition S3 *For a tripartite population consisting of fractions f_- and f_+ with negative and positive attitude, i.e. $a_- < 0$, $a_+ > 0$, and a neutral fraction f_0 with neutral attitude, $a_0 = 0$, with no a priori preference for either sideward, such that $f_- + f_+ + f_0 = 1$ the results obtained for a bipartite population are modified as follows:*
(1) the aggregate opinions are normally distributed according to a normal distribution with mean

$$\mu(B) = f_- \cdot \mu_-(B) + f_0 \cdot \mu_0(B) + f_+ \cdot \mu_+(B)$$

and variance

$$\sigma^2(N, B) = \frac{1}{N}\left(f_+\sigma_+^2(B) + f_0\sigma_0^2(B) + f_-\sigma_-^2(B)\right).$$

(2) The polarization index in a tripartite population assumes the form

$$q(B) = 2(\mu_+(B) - \mu(B))f_+ - 2(\mu_-(B) - \mu(B))f_-,$$

and simplifies to

$$q(B) = (1 - f_0) \cdot \tilde{q}(B)$$

in the symmetric case where $f_+ = f_- = (1 - f_0)/2$ and $\tilde{q}(B)$ denotes the polarization of a bipartite symmetric population with $f_+ = f_- = 1/2$.

Proof of Proposition S3: For a tripartite population with attitudes $a_- < 0$, $a_0 = 0$ and $a_+ > 0$ we find

$$\mu_+(B) = \frac{1}{1 + e^{-\beta(a+B)}}, \quad \mu_0(B) = \frac{1}{1 + e^{-\beta B}}, \quad \mu_-(B) = \frac{1}{1 + e^{\beta(a-B)}}$$

for the individual population segments as well as

$$\sigma_\pm^2(N_\pm, B) = \frac{\sigma_\pm^2(B)}{N_\pm} = \frac{f_\pm \cdot \sigma_\pm^2(B)}{N} \text{ and } \sigma_0^2(N_0, B) = \frac{\sigma_0^2(B)}{N_0} = \frac{f_0 \cdot \sigma_0^2(B)}{N}$$

where

$$\sigma_\pm^2(B) = (1 - \mu_\pm(B)) \cdot \mu_\pm(B) \text{ and } \sigma_0^2(B) = (1 - \mu_0(B)) \cdot \mu_0(B)$$

For the average opinion and variance then follows

$$\mu(B) = f_-\mu_-(B) + f_0\mu_0(B) + f_+\mu_+(B)$$

and

$$\sigma^2(N, B) = \frac{1}{N}\left(f_+\sigma_+^2(B) + f_0\sigma_0^2(B) + f_-\sigma_-^2(B)\right)$$

The polarization index is defined as

$$q(B) = 2(\mu_+(B) - \mu(B))f_+ - 2(\mu_-(B) - \mu(B))f_-$$

For a balanced scenario, where

$$f_- = (1 - f_0)/2 = f_+$$

the term $\mu(B)$ drops out of the expression for $q(B)$ and the polarization index simplifies to

$$q(B) = (1 - f_0) \cdot (\mu_+(B) - \mu_-(B)),$$

or

$$q(B) = (1 - f_0) \cdot \left(\frac{1}{1 + e^{-\beta(a+B)}} - \frac{1}{1 + e^{\beta(a-B)}} \right).$$

A comparison with the result for the bipartite case, discussed in Proposition 4, proves the proposition. **QED**