Supplemental Appendix
Coarse Memory and Plausible Narratives

Francesco Bilotta and Giacomo Manferdini

A General Narrative Design

This Appendix studies a general version of the Narrative Design problem. In particular, we allow for any finite policy set A, any outcome measure space Y and any bounded measurable utility function $u : Y \to \mathbb{R}$. In this case, the Plausibility constraint rewrites

$$\mathcal{M}(\alpha, \upsilon) = \left\{ \mu : A \to \Delta(Y) \;\middle|\; \upsilon = \sum_{a \in A} \alpha(a)\mu_a \right\}. \tag{1}$$

and, fixed a policy a, the politician's problem remains

$$V_a(\alpha, \upsilon) = \max_{\mu \in \mathcal{M}(\alpha,\upsilon)} \mathbb{E}_{\mu_a}[u(y)]. \qquad \text{(General Narrative Design)}$$

That is, the politician chooses a plausible narrative to maximize the voter's anticipatory utility from policy a. Then, Proposition 1 can be generalized to the following.

Proposition A.1 *Fix a coarse memory $(\alpha, \upsilon) \in \Delta(A) \times \Delta(Y)$ and $a \in A$ with $\alpha(a) \in (0, 1)$. Define the utility threshold $\hat{u} \equiv \inf \left\{ r \in \mathbb{R} \;\middle|\; \upsilon\big(\{y \in Y :\; u(y) > r\}\big) \leq \alpha(a) \right\}$. Let $c \in [0, 1]$ satisfy $\alpha(a) = \upsilon\big(\{u > \hat{u}\}\big) + c\,\upsilon\big(\{u = \hat{u}\}\big)$, with the convention $c = 0$ when $\upsilon(\{u = \hat{u}\}) = 0$. The General Narrative Design problem is solved by any $\hat{\mu} \in \mathcal{M}(\alpha, \upsilon)$ such that $\hat{\mu}_a$ has Radon–Nikodym derivative*

$$\frac{d\hat{\mu}_a}{d\upsilon} = \frac{1}{\alpha(a)}\Big(\mathbf{1}_{\{u>\hat{u}\}} + c\,\mathbf{1}_{\{u=\hat{u}\}}\Big).$$

The corresponding value is

$$V_a(\alpha, \upsilon) = \frac{1}{\alpha(a)} \int_Y u(y)\Big(\mathbf{1}_{\{u(y)>\hat{u}\}} + c\,\mathbf{1}_{\{u(y)=\hat{u}\}}\Big)\, d\upsilon(y).$$

The Proposition states that the optimal narrative assigns to the commitment policy a exactly the top $\alpha(a)$-mass of outcomes according to the utility ranking induced by u (with tie-breaking when υ has atoms). Equivalently, $\hat{\mu}_a$ is obtained by conditioning υ on a superlevel set of u of υ-mass $\alpha(a)$. Two implications are worth noting. First, the construction depends on u only through its level sets, so any strictly increasing transformation of u induces the same optimal $\hat{\mu}_a$, showing that our solution is independent of risk attitudes. Second, when

$|A| > 2$, the politician is indifferent over how the remaining $(1 - \alpha(a))$ mass is distributed across policies $a' \neq a$, hence, the allocation across individual $a' \neq a$ will not be unique. Nonetheless, the distribution of outcomes conditional on the event "not a" is uniquely pinned down. For any narrative μ, define the average distribution associated with policies different from a,

$$\mu_{\neg a} \;\equiv\; \frac{1}{1-\alpha(a)} \sum_{a' \neq a} \alpha(a')\mu_{a'} \;\in \Delta(Y).$$

Then the constraint in Equation (1) can be equivalently written as

$$\alpha(a)\mu_a + (1-\alpha(a))\mu_{\neg a} = \upsilon, \tag{2}$$

and we have the following consequence.

Corollary A.1 *Under the conditions of Proposition A.1, in any optimal narrative $\hat{\mu}$ the distribution $\hat{\mu}_{\neg a}$ is uniquely pinned down and satisfies*

$$\frac{d\hat{\mu}_{\neg a}}{d\upsilon} = \frac{1}{1-\alpha(a)}\Big(\mathbf{1}_{\{u<\hat{u}\}} + (1-c)\,\mathbf{1}_{\{u=\hat{u}\}}\Big).$$

In particular, all optimal narratives induce the same pair $(\hat{\mu}_a, \hat{\mu}_{\neg a})$, which we call a *sufficient representation.*

Proposition A.1 follows from the so-called *bathtub principle* (e.g., Theorem 1.14 in Lieb and Loss, 2001), which asserts that to maximize the integral of a function (here, u) over all sets of fixed measure (here, $\alpha(a)$) one must choose an upper level set of the function: thinking the graph of the function as a bathtub, one "fills it from top layers down." This is closely related to the concept of non-decreasing rearrangement of a function and the Hardy-Littlewood inequality. Corollary A.1 follows immediately from Equation (2) by taking Radon–Nikodym derivatives with respect to υ.

A.1 Comparative Statics

We next study comparative statics of General Narrative Design with respect to policy frequencies α and the true model μ^*. In contrast to the binary case, an increase in the frequency of one policy mechanically forces the frequencies of the others to fall only *in the aggregate*. Many patterns of adjustment are therefore possible. To preserve our sharp comparative statics, we focus on variations in which the frequency of the committed policy changes *against* all others: whenever $\alpha(a)$ rises, each $\alpha(a')$ for $a' \neq a$ weakly falls. This restriction captures the idea that the frequency of policy a increases (or decreases) independently of the relative composition among the remaining policies. It is also the relevant variation in our application, where α evolves through elections with a single winner. Fix a committed policy $a \in A$. For $\alpha, \alpha' \in \Delta(A)$, say that α' *a-majorizes* α if

$$\alpha'(a) \geq \alpha(a) \quad \text{and} \quad \alpha'(a') \leq \alpha(a') \text{ for all } a' \neq a.$$

For models, say that $\mu^{*\prime}$ is *weakly more productive than* μ^* if for every policy $a \in A$ and every utility threshold $r \in \mathbb{R}$,

$$\mu_a^{*\prime}\big(\{y : u(y) \geq r\}\big) \geq \mu_a^*\big(\{y : u(y) \geq r\}\big).$$

When u is increasing, this condition reduces to policy-wise first-order stochastic dominance, as in the main text.

Proposition A.2 *Fix $a \in A$. The following comparative statics holds:*

1. *Fix μ^*. If α' a-majorizes α, then $V_a\big(\alpha', \upsilon(\alpha', \mu^*)\big) \leq V_a\big(\alpha, \upsilon(\alpha, \mu^*)\big)$.*
2. *Fix α. If $\mu^{*\prime}$ is weakly more productive than μ^*, then $V_a\big(\alpha, \upsilon(\alpha, \mu^{*\prime})\big) \geq V_a\big(\alpha, \upsilon(\alpha, \mu^*)\big)$.*

A.2 Connection to Partial Identification

Our plausibility restriction admits a natural interpretation through the lens of partial identification econometrics. In that literature, observed information and maintained restrictions typically pin down a *set* of admissible structural parameters for a data generating process rather than a unique one. In our model, the voter observes only policy frequencies α and the marginal outcome distribution υ, hence, the set of plausible narratives $\mathcal{M}(\alpha, \upsilon)$ can be interpreted as the partially identified set for the true correlation structure. In particular, the mixture decomposition

$$\upsilon = \alpha(a)\mu_a + (1 - \alpha(a))\mu_{\neg a}. \tag{3}$$

is formally identical to the mixture structure studied in Horowitz and Manski (1995), where an observed distribution is expressed as a known-weight mixture of a target distribution and an unrestricted residual component.

In Proposition 4 and Corollary 4.1, Horowitz and Manski characterize the *sharp bounds* on monotone functionals—such as the mean of an increasing function—over the set of distributions consistent with such a mixture. Our General Narrative Design problem extends this result in two directions. First, it computes the sharp upper bound of a non-monotone welfare functional. Moreover, our solution makes the bound constructive: it characterizes the extremal distribution $\hat{\mu}_a$ that attains it by concentrating probability mass on the highest-utility outcomes allowed by Equation (3).

Additionally, the comparative statics and dynamics of our model parallel those studied in the social learning problem of Manski (2004), where identification regions shrink as an action is taken more often, because additional observations tighten the feasible set of outcome distributions. However, our coarse-memory assumption prevents the joint distribution from ever being fully recovered, even with large samples, and the tightness of the identified sets varies with the endogenous implementation frequencies.

A.3 Connection with Optimal Transport

The narrative design problem can also be viewed as an optimal transport problem. Let $\Gamma(\alpha, \upsilon)$ denote the set of *couplings* between α and υ, i.e., joint distributions $\gamma \in \Delta(A \times Y)$ with marginals α and υ.

There is a canonical correspondence between plausible narratives and couplings. Given $\mu \in \mathcal{M}(\alpha, \upsilon)$, define the joint distribution γ^μ by

$$\gamma^\mu(\{a\} \times B) = \alpha(a)\mu_a(B) \qquad \text{for all } a \in A,\ B \subseteq Y \text{ measurable.}$$

Conversely, given $\gamma \in \Gamma(\alpha, \upsilon)$, define μ^γ by

$$\mu_a^\gamma(B) = \frac{\gamma(\{a\} \times B)}{\alpha(a)} \qquad \text{for all } a \in A,\ B \subseteq Y \text{ measurable,}$$

with an arbitrary extension on actions a with $\alpha(a) = 0$. This identifies $\mathcal{M}(\alpha, \upsilon)$ with $\Gamma(\alpha, \upsilon)$.

Under this correspondence, the objective in General Narrative Design becomes

$$\mathbb{E}_{\mu_a}[u(y)] = \frac{1}{\alpha(a)} \int_{A \times Y} u(y) \mathbf{1}_{\{a'=a\}}\, d\gamma(a', y).$$

Define the surplus

$$\sigma_a(a', y) = \frac{u(y)}{\alpha(a)} \mathbf{1}_{\{a'=a\}}.$$

Then General Narrative Design is equivalent to the optimal transport problem

$$\max_{\gamma \in \Gamma(\alpha, \upsilon)} \int_{A \times Y} \sigma_a(a', y)\, d\gamma(a', y). \tag{4}$$

Restrict attention to the case when $Y \subseteq \mathbb{R}$ and u is increasing and embed A into $\mathbb{R}$ via some injection $\iota : A \hookrightarrow \mathbb{R}$ with $\iota(a) = \max \iota(A)$. In this case, viewed as a function on $\mathbb{R}^2$, σ_a

is supermodular: for any $x', x'' \in \iota(A)$ and $y', y'' \in \mathbb{R}$,

$$\sigma_a(\max\{x', x''\}, \max\{y', y''\}) + \sigma_a(\min\{x', x''\}, \min\{y', y''\}) \;\geq\; \sigma_a(x', y') + \sigma_a(x'', y'').$$

Indeed, when both a' and a'' are different from a, both the RHS and the LHS equal 0. When both are equal to a, both the RHS and the LHS are equal to $\sigma_a(a, y') + \sigma_a(a, y'')$. When only one of the two is equal to a, say, without loss a', the inequality reduces to

$$\sigma_a(a', \max\{y', y''\}) \geq \sigma_a(a', y')$$

which holds since u is increasing. Therefore, applying Theorem 4.3.(i) from Galichon (2018) we obtain that the solution $\hat{\gamma}_a$ to Equation (4) must be the *comonotone coupling* between α and υ. In other words, there exists a uniform random variable $U \sim \mathcal{U}([0,1])$ such that the random vector $(\alpha^{-1}(U), \upsilon^{-1}(U))$ is distributed according to $\hat{\gamma}_a$. This alternative characterization of the optimal narrative shows how it realizes *positive assortative matching* between the policy and the outcome distribution. While we do not pursue the optimal transport direction further here, it provides a useful alternative perspective and suggests natural generalizations of the model.

B Alternative Voter's Decision Rule

In the baseline game, the voter evaluates policy g using G's narrative and policy b using B's narrative. A natural alternative timing separates narrative selection from voting.

In Stage 1, the voter adopts the most hopeful narrative, anticipating that politicians implement their commitments. That is, she adopts μ^G if

$$\mathbb{E}_{\mu^G(g)}[u(Y)] \;\geq\; \mathbb{E}_{\mu^B(b)}[u(Y)], \tag{5}$$

and adopts μ^B otherwise.

In Stage 2, she evaluates both candidates using the adopted narrative. Let μ^W denote the narrative selected in Stage 1 and w its associated policy. The voter elects W if

$$\mathbb{E}_{\mu^W(w)}[u(Y)] \;\geq\; \mathbb{E}_{\mu^W(\neg w)}[u(Y)] + \varphi,$$

and elects $\neg W$ otherwise.

Under this timing, it remains optimal for each politician to announce the solution to the narrative design problem. Let $\hat{\mu}^G$ and $\hat{\mu}^B$ denote the corresponding optimal narratives. By construction,

$$\mathbb{E}_{\hat{\mu}^{W}(\neg w)}[u(Y)] \;\le\; \mathbb{E}_{\hat{\mu}^{\neg W}(\neg w)}[u(Y)],$$

since $\hat{\mu}^{\neg W}$ maximizes the perceived value of policy $\neg w$. Thus, the narrative selected in Stage 1 always assigns a weakly lower value to the opponent's policy than the opponent would assign to himself. As a consequence, the winner of the narrative stage enjoys a (weakly) larger advantage in the voting stage than in the baseline model. Intuitively, once a narrative prevails, it governs the evaluation of both policies, and therefore correlates the opponent's policy with worse outcomes than the opponent's own narrative would have allowed.

Hence, at any coarse memory where G (respectively B) had a higher probability of winning in the baseline model, his probability of winning is even higher under the modified timing. The function $w^G(\alpha)$ becomes steeper around its fixed point at $\alpha = \frac{1}{2}$, amplifying the impact of narrative advantage. In the dynamic extension, this steeper response translates into more persistent incumbency spells and, correspondingly, longer policy cycles. The qualitative properties of convergence and polarization remain unchanged.

C Polarization under Bounded f-Divergences

This appendix shows that Proposition 3 is robust to a large class of bounded statistical divergences. For the sake of this extension, we work on the nondegenerate domain $\alpha \in (0,1)$ and we assume that μ^* is continuous, so that $\upsilon(\alpha, \mu^*) = \alpha\mu_g^* + (1-\alpha)\mu_b^*$ is continuous, too.

We call a convex function $f : (0,\infty) \to \mathbb{R}$ such that $f(1) = 0$, $0 \in \partial f(1)$, admitting finite limits $f(0) = \lim_{t\downarrow 0} f(t)$ and $f^\infty = \lim_{t\to\infty} f(t)/t$ a *bounded generator*. Let $\kappa_f := f(0) + f^\infty > 0$. For $\lambda, \lambda' \in \Delta(Y)$, let m be any measure dominating both, and write $p = d\lambda/dm$ and $q = d\lambda'/dm$. Any bounded generator f defines a *bounded f-divergence* as

$$D_f(\lambda\|\lambda') = \int_Y q\, f\left(\frac{p}{q}\right) dm,$$

with the standard convention that the integrand is $qf(0)$ when $p = 0 < q$, is pf^∞ when $q = 0 < p$, and is zero when $p = q = 0$.

This class of divergences includes several well-known examples, for different choices of f.

- Total variation: if $f(t) = \frac{1}{2}|t-1|$, then $D_f(P\|Q) = d^{TV}(P,Q)$.
- Squared Hellinger distance: if $f(t) = \frac{1}{2}(\sqrt{t}-1)^2$, then $D_f(P\|Q) = H^2(P,Q)$, where $H^2(P,Q) = 1 - \int \sqrt{pq}\, dm$.

- Jensen–Shannon divergence: if $f(t) = \frac{1}{2}\left[t \log \frac{2t}{1+t} + \log \frac{2}{1+t}\right]$. Then $D_f(P\|Q) = JS(P,Q)$, where $JS(P,Q) = \frac{1}{2}KL(P\|M) + \frac{1}{2}KL(Q\|M)$ and $M = (P+Q)/2$.
- Triangular discrimination: if $f(t) = \frac{(t-1)^2}{t+1}$, then $D_f(P\|Q) = \int (p-q)^2/(p+q)\, dm$.

We consider the normalized symmetric version of f-divergences,

$$d^f(\lambda, \lambda') = \frac{D_f(\lambda\|\lambda') + D_f(\lambda'\|\lambda)}{2\kappa_f},$$

and extend it to narratives $\mu, \mu' \in \mathcal{M}(\alpha, \upsilon)$ as

$$d^{\mathcal{M},f}(\mu, \mu') = \frac{1}{2}\left[d^f(\mu_g, \mu'_g) + d^f(\mu_b, \mu'_b)\right].$$

Our main robustness result is the following.

Proposition C.1 *Fix $\alpha \in (0,1)$ and a continuous true model μ^*. For every bounded generator f, the equilibrium narratives $(\hat{\mu}^G, \hat{\mu}^B)$ maximize $d^{\mathcal{M},f}(\mu, \mu')$ over $\mathcal{M}(\alpha, \upsilon) \times \mathcal{M}(\alpha, \upsilon)$. Moreover, $d^{\mathcal{M},f}(\hat{\mu}^G, \hat{\mu}^B)$ is maximized at $\alpha = \frac{1}{2}$.*

D Additional Proofs

Proof of Proposition A.1. Fix $(\alpha, \upsilon) \in \Delta(A) \times \Delta(Y)$ and a policy $a \in A$ with $\alpha(a) \in (0,1)$. Write $\mathcal{M}(\alpha, \upsilon)$ as in Equation (1). We first show two simple facts.

Fact 1. Let $\mu \in \mathcal{M}(\alpha, \upsilon)$ and fix $a \in A$ with $\alpha(a) > 0$. Then $\mu_a \ll \upsilon$.

Proof. Suppose by contradiction that there exists a measurable set $B \subseteq Y$ such that $\upsilon(B) = 0$ but $\mu_a(B) > 0$. Then, using $\upsilon = \sum_{a' \in A} \alpha(a')\mu_{a'}$ (plausibility, as in Equation (1)),

$$\upsilon(B) = \sum_{a' \in A} \alpha(a')\mu_{a'}(B) \;\geq\; \alpha(a)\mu_a(B) \;>\; 0,$$

a contradiction. ■

Fact 2. For any $\mu \in \mathcal{M}(\alpha, \upsilon)$ and any $a \in A$ with $\alpha(a) > 0$,

$$\frac{d\mu_a}{d\upsilon}(y) \;\leq\; \frac{1}{\alpha(a)} \qquad \upsilon\text{-a.e.}$$

Proof. Let $h \equiv \frac{d\mu_a}{d\upsilon}$, which exists by Fact 1 and the Radon–Nikodym theorem. If the claim failed, there would exist a measurable $B \subseteq Y$ with $\upsilon(B) > 0$ such that $h(y) > \frac{1}{\alpha(a)}$ for all $y \in B$. Then

$$\mu_a(B) = \int_B h\, d\upsilon \;>\; \frac{1}{\alpha(a)}\upsilon(B),$$

so $\alpha(a)\mu_a(B) > \upsilon(B)$. But by plausibility (Equation (1)),

$$\upsilon(B) = \sum_{a' \in A} \alpha(a')\mu_{a'}(B) \ \geq \ \alpha(a)\mu_a(B),$$

a contradiction. ■

Fix $a \in A$. For any $\mu \in \mathcal{M}(\alpha, \upsilon)$ define

$$g(y) \ \equiv \ \alpha(a)\,\frac{d\mu_a}{d\upsilon}(y).$$

By Fact 2, $0 \leq g \leq 1$ υ-a.e. Moreover,

$$\int_Y g\,d\upsilon = \alpha(a)\int_Y \frac{d\mu_a}{d\upsilon}\,d\upsilon = \alpha(a)\mu_a(Y) = \alpha(a).$$

Hence g belongs to the class

$$C(\alpha(a), \upsilon) \ \equiv \ \Big\{g: \ 0 \leq g \leq 1 \ \upsilon\text{-a.e.}, \ \int_Y g\,d\upsilon = \alpha(a)\Big\}.$$

Conversely, given any $g \in C(\alpha(a), \upsilon)$, define a probability measure μ_a^g by

$$\mu_a^g(B) \ \equiv \ \frac{1}{\alpha(a)}\int_B g\,d\upsilon, \qquad B \subseteq Y \text{ measurable.}$$

Then $\mu_a^g \in \Delta(Y)$. Define the (nonnegative) residual measure

$$\rho^g \ \equiv \ \upsilon - \alpha(a)\mu_a^g, \qquad \text{so that} \qquad \rho^g(B) = \int_B (1-g)\,d\upsilon \ \geq \ 0,$$

and note that $\rho^g(Y) = 1 - \alpha(a)$. Choose any assignment of the residual mass across the remaining policies, for instance set $\mu_{a'}^g \equiv \rho^g/(1-\alpha(a))$ for every $a' \neq a$. Then $\mu^g \equiv (\mu_{a'}^g)_{a' \in A}$ belongs to $\mathcal{M}(\alpha, \upsilon)$ and satisfies

$$\alpha(a)\frac{d\mu_a^g}{d\upsilon} = g.$$

Therefore, maximizing over $\mu \in \mathcal{M}(\alpha, \upsilon)$ is equivalent to maximizing over $g \in C(\alpha(a), \upsilon)$, and

$$\begin{aligned} V_a(\alpha, \upsilon) &= \max_{\mu \in \mathcal{M}(\alpha,\upsilon)} \int_Y u(y)\,d\mu_a(y) \\ &= \max_{g \in C(\alpha(a),\upsilon)} \ \frac{1}{\alpha(a)}\int_Y u(y)\,g(y)\,d\upsilon(y). \end{aligned}$$

Since υ is a finite measure and u is bounded measurable, we may apply the bathtub principle (e.g. Lieb and Loss, 2001, Theorem 1.14) to the minimization of $\int(-u)g\,d\upsilon$ over $g \in C(\alpha(a), \upsilon)$. Equivalently, it yields a maximizer of $\int ug\,d\upsilon$ over the same constraint set.

Let

$$\hat{u} \;\equiv\; \inf\Big\{r \in \mathbb{R} \;\Big|\; \upsilon(\{u > r\}) \le \alpha(a)\Big\}.$$

Choose $c \in [0,1]$ such that

$$\alpha(a) = \upsilon(\{u > \hat{u}\}) + c\,\upsilon(\{u = \hat{u}\}),$$

with the convention $c = 0$ when $\upsilon(\{u = \hat{u}\}) = 0$. Then the bathtub principle implies that an optimizer g^* is given υ-a.e. by

$$g^*(y) = \mathbf{1}_{\{u(y)>\hat{u}\}} + c\,\mathbf{1}_{\{u(y)=\hat{u}\}}.$$

Therefore an optimal narrative can be chosen so that

$$\alpha(a)\frac{d\hat{\mu}_a}{d\upsilon} = g^*, \qquad \text{i.e.} \qquad \frac{d\hat{\mu}_a}{d\upsilon} = \frac{1}{\alpha(a)}\Big(\mathbf{1}_{\{u>\hat{u}\}} + c\,\mathbf{1}_{\{u=\hat{u}\}}\Big),$$

which is exactly the formula in the statement. Plugging this into the objective yields the value

$$V_a(\alpha, \upsilon) = \int_Y u(y)\,d\hat{\mu}_a(y) = \frac{1}{\alpha(a)}\int_Y u(y)\Big(\mathbf{1}_{\{u(y)>\hat{u}\}} + c\,\mathbf{1}_{\{u(y)=\hat{u}\}}\Big)\,d\upsilon(y).$$

This completes the proof. □

Proof of Corollary A.1. Under the conditions of the proposition, write plausibility in the two-component form of Equation (2):

$$\alpha(a)\hat{\mu}_a + (1-\alpha(a))\hat{\mu}_{\neg a} = \upsilon.$$

By Fact 1 above, both $\hat{\mu}_a$ and $\hat{\mu}_{\neg a}$ are absolutely continuous with respect to υ. Taking Radon–Nikodym derivatives with respect to υ gives υ-a.e.

$$\alpha(a)\frac{d\hat{\mu}_a}{d\upsilon} + (1-\alpha(a))\frac{d\hat{\mu}_{\neg a}}{d\upsilon} = 1.$$

Substituting the expression for $\frac{d\hat{\mu}_a}{d\upsilon}$ from Proposition A.1,

$$(1-\alpha(a))\frac{d\hat{\mu}_{\neg a}}{d\upsilon} = 1 - \Big(\mathbf{1}_{\{u>\hat{u}\}} + c\,\mathbf{1}_{\{u=\hat{u}\}}\Big) = \mathbf{1}_{\{u<\hat{u}\}} + (1-c)\,\mathbf{1}_{\{u=\hat{u}\}},$$

so

$$\frac{d\hat{\mu}_{\neg a}}{dv} = \frac{1}{1-\alpha(a)}\Big(\mathbf{1}_{\{u<\hat{u}\}} + (1-c)\,\mathbf{1}_{\{u=\hat{u}\}}\Big),$$

as claimed. Uniqueness of $\hat{\mu}_{\neg a}$ follows because 2 pins it down once $\hat{\mu}_a$ is fixed. □

Proof of Proposition A.2. Fix a commitment policy $a \in A$. For any (α, υ) with $\alpha(a) \in (0,1)$, define the set of *a-feasible conditional distributions*

$$\mathcal{F}_a(\alpha,\upsilon) \;\equiv\; \Big\{\lambda \in \Delta(Y) \;:\; \alpha(a)\lambda \le \upsilon \text{ as measures}\Big\},$$

i.e. $\alpha(a)\lambda(B) \le \upsilon(B)$ for every measurable $B \subseteq Y$.

Lemma D.1 *For any* (α,υ) *with* $\alpha(a) \in (0,1)$,

$$V_a(\alpha,\upsilon) = \sup_{\lambda\in\mathcal{F}_a(\alpha,\upsilon)} \int_Y u(y)\,d\lambda(y).$$

Proof. If $\mu \in \mathcal{M}(\alpha,\upsilon)$, then by Equation (1), $\upsilon = \sum_{a'\in A}\alpha(a')\mu_{a'} \ge \alpha(a)\mu_a$, hence $\mu_a \in \mathcal{F}_a(\alpha,\upsilon)$, which shows $V_a(\alpha,\upsilon) \le \sup_{\lambda\in\mathcal{F}_a(\alpha,\upsilon)}\int u\,d\lambda$. Conversely, take any $\lambda \in \mathcal{F}_a(\alpha,\upsilon)$ and set the residual measure $\rho \equiv \upsilon - \alpha(a)\lambda$, which is nonnegative by definition and has total mass $1-\alpha(a)$. Define $\mu_a = \lambda$ and, for each $a' \ne a$, set $\mu_{a'} \equiv \rho/(1-\alpha(a))$. Then $\mu \in \mathcal{M}(\alpha,\upsilon)$ and achieves $\int u\,d\lambda$, so equality holds. ■

We now prove the two parts.

(a) Monotonicity in a-majorization (fixed μ^*). Fix μ^*. For any $\alpha \in \Delta(A)$ define the induced outcome marginal

$$\upsilon(\alpha,\mu^*) \;\equiv\; \sum_{a'\in A}\alpha(a')\,\mu^*_{a'} \in \Delta(Y).$$

Assume that α' a-majorizes α, i.e. $\alpha'(a) \ge \alpha(a)$ and $\alpha'(a') \le \alpha(a')$ for all $a' \ne a$. We claim that

$$\frac{1}{\alpha'(a)}\,\upsilon(\alpha',\mu^*) \;\le\; \frac{1}{\alpha(a)}\,\upsilon(\alpha,\mu^*) \qquad \text{as measures.} \tag{6}$$

Indeed, for any measurable $B \subseteq Y$,

$$\begin{aligned}\frac{\upsilon(\alpha',\mu^*)(B)}{\alpha'(a)} &= \mu^*_a(B) + \sum_{a'\ne a}\frac{\alpha'(a')}{\alpha'(a)}\,\mu^*_{a'}(B)\\ &\le \mu^*_a(B) + \sum_{a'\ne a}\frac{\alpha(a')}{\alpha(a)}\,\mu^*_{a'}(B) = \frac{\upsilon(\alpha,\mu^*)(B)}{\alpha(a)},\end{aligned}$$

where the inequality uses $\alpha'(a')/\alpha'(a) \le \alpha(a')/\alpha(a)$ for each $a' \ne a$ and $\mu^*_{a'}(B) \ge 0$.

Now let $\lambda \in \mathcal{F}_a(\alpha', \upsilon(\alpha', \mu^*))$. By definition, $\lambda(B) \le \upsilon(\alpha', \mu^*)(B)/\alpha'(a)$ for all B. Combining with Equation (6) gives $\lambda(B) \le \upsilon(\alpha, \mu^*)(B)/\alpha(a)$ for all B, hence

$$\mathcal{F}_a(\alpha', \upsilon(\alpha', \mu^*)) \;\subseteq\; \mathcal{F}_a(\alpha, \upsilon(\alpha, \mu^*)).$$

Applying Lemma D.1 and taking suprema of the same linear functional $\lambda \mapsto \int u\, d\lambda$ over nested feasible sets yields

$$V_a(\alpha', \upsilon(\alpha', \mu^*)) \;\le\; V_a(\alpha, \upsilon(\alpha, \mu^*)).$$

(b) Monotonicity in productivity (fixed α). Fix α. Let $\mu^{*'}$ be weakly more productive than μ^*, i.e. for every $a' \in A$ and every $r \in \mathbb{R}$,

$$\mu^{*'}_{a'}(\{y : \; u(y) \ge r\}) \;\ge\; \mu^*_{a'}(\{y : \; u(y) \ge r\}).$$

Let $\upsilon \equiv \upsilon(\alpha, \mu^*)$ and $\upsilon' \equiv \upsilon(\alpha, \mu^{*'})$. Summing the above inequalities with weights $\alpha(a')$ gives, for every r,

$$\upsilon'(\{u \ge r\}) \;\ge\; \upsilon(\{u \ge r\}). \tag{7}$$

Let $U = u(Y)$ when $Y \sim \upsilon$ and $U' = u(Y')$ when $Y' \sim \upsilon'$. Then Equation (7) is exactly $\mathbb{P}(U' \ge r) \ge \mathbb{P}(U \ge r)$ for all r, i.e. U' first-order stochastically dominates U.

Let Q and Q' be the generalized quantile functions of U and U': $Q(p) = \inf\{r : \; \mathbb{P}(U \le r) \ge p\}$ and similarly for Q'. FOSD implies $Q'(p) \ge Q(p)$ for all $p \in [0,1]$. By Proposition A.1, the value $V_a(\alpha, \upsilon)$ is the average of the top $\alpha(a)$-mass of realizations of U (with the conventional tie-handling when υ has atoms on $\{u = \hat{u}\}$). Equivalently,

$$V_a(\alpha, \upsilon) = \frac{1}{\alpha(a)} \int_{1-\alpha(a)}^{1} Q(p)\, dp, \qquad V_a(\alpha, \upsilon') = \frac{1}{\alpha(a)} \int_{1-\alpha(a)}^{1} Q'(p)\, dp.$$

Since $Q'(p) \ge Q(p)$ pointwise, integrating over $[1 - \alpha(a), 1]$ yields

$$V_a(\alpha, \upsilon') \;\ge\; V_a(\alpha, \upsilon),$$

which is the desired conclusion. □

Proof of Proposition C.1. The proof has two steps. First, plausibility bounds how different two feasible policy-contingent distributions can be. Second, Proposition 1 shows that the equilibrium narratives attain this bound because they assign each policy to opposite tails of the same marginal υ.

Fix $\alpha \in (0,1)$ and write $\upsilon = \upsilon(\alpha, \mu^*)$. Since μ^* is continuous, υ is atomless. Take arbitrary $\mu, \mu' \in \mathcal{M}(\alpha, \upsilon)$. Plausibility gives

$$\alpha\mu_g + (1-\alpha)\mu_b = \upsilon \qquad \text{and} \qquad \alpha\mu'_g + (1-\alpha)\mu'_b = \upsilon.$$

Hence $\mu_g, \mu'_g \ll \upsilon$, with $d\mu_g/d\upsilon, d\mu'_g/d\upsilon \leq 1/\alpha$, and $\mu_b, \mu'_b \ll \upsilon$, with $d\mu_b/d\upsilon, d\mu'_b/d\upsilon \leq 1/(1-\alpha)$. Consider first the total variation distance,

$$d^{TV}(\mu_g, \mu'_g) = 1 - \int_Y \min\{d\mu_g/d\upsilon, d\mu'_g/d\upsilon\}\, d\upsilon \leq \min\left\{1, \frac{1-\alpha}{\alpha}\right\}.$$

Indeed, if $\alpha > 1/2$, then $\int \min\{d\mu_g/d\upsilon, d\mu'_g/d\upsilon\}\, d\upsilon = 2 - \int \max\{d\mu_g/d\upsilon, d\mu'_g/d\upsilon\}\, d\upsilon \geq 2 - 1/\alpha$; if $\alpha \leq 1/2$, the bound is the trivial one. The same argument applied to the policy-b components gives

$$d^{TV}(\mu_b, \mu'_b) \leq \min\left\{1, \frac{\alpha}{1-\alpha}\right\}.$$

Next, observe that bounded f-divergences are bounded above by total variation after the normalization used in the proposition. Let $P, Q \in \Delta(Y)$, and let m dominate both, with densities $p = dP/dm$ and $q = dQ/dm$. By convexity, $f(1) = 0$, and the endpoint bounds $f(0) < \infty$ and $f^\infty < \infty$, one has $f(t) \leq f(0)(1-t)$ for $t \leq 1$, and $f(t) \leq f^\infty(t-1)$ for $t \geq 1$. Thus

$$D_f(P\|Q) \leq f(0)\int (q-p)_+\, dm + f^\infty \int (p-q)_+\, dm = \kappa_f d^{TV}(P,Q).$$

The same inequality holds after interchanging P and Q, so $d^f(P,Q) \leq d^{TV}(P,Q)$. Combining this with the two total-variation bounds above yields, for every $\mu, \mu' \in \mathcal{M}(\alpha, \upsilon)$,

$$d^{\mathcal{M},f}(\mu, \mu') \leq \frac{1}{2}\left[\min\left\{1, \frac{1-\alpha}{\alpha}\right\} + \min\left\{1, \frac{\alpha}{1-\alpha}\right\}\right]. \tag{8}$$

It remains to show that the equilibrium narratives attain the bound in Equation (8). By Proposition 1, politician G's optimal narrative assigns policy g to the upper α-tail of υ and policy b to the lower $(1-\alpha)$-tail of υ. By the symmetric statement for politician B, his optimal narrative assigns policy g to the lower α-tail and policy b to the upper $(1-\alpha)$-tail.

Consider first the policy-g components. Since υ is atomless, the upper and lower α-tails have υ-mass α, and their overlap has υ-mass $\max\{0, 2\alpha - 1\}$. On this overlap, $\hat{\mu}_g^G$ and $\hat{\mu}_g^B$ have the same density $1/\alpha$ with respect to υ, so the likelihood ratio is one and the contribution to the f-divergence is zero. Off the overlap, their masses are supported on disjoint tail regions, so the only contribution comes from the endpoint values $f(0)$ and f^∞.

Hence, after normalization by $\kappa_f = f(0) + f^\infty$,

$$d^f(\hat{\mu}_g^G, \hat{\mu}_g^B) = 1 - \frac{\max\{0, 2\alpha - 1\}}{\alpha} = \min\left\{1, \frac{1-\alpha}{\alpha}\right\}.$$

The same argument applies to the policy-b components. Here the relevant tails have υ-mass $1-\alpha$, and their overlap has υ-mass $\max\{0, 1-2\alpha\}$. Therefore

$$d^f(\hat{\mu}_b^G, \hat{\mu}_b^B) = 1 - \frac{\max\{0, 1 - 2\alpha\}}{1-\alpha} = \min\left\{1, \frac{\alpha}{1-\alpha}\right\}.$$

It follows that

$$d^{\mathcal{M},f}(\hat{\mu}^G, \hat{\mu}^B) = \frac{1}{2}\left[\min\left\{1, \frac{1-\alpha}{\alpha}\right\} + \min\left\{1, \frac{\alpha}{1-\alpha}\right\}\right].$$

Together with Equation (8), this proves that $(\hat{\mu}^G, \hat{\mu}^B)$ maximizes $d^{\mathcal{M},f}$ over $\mathcal{M}(\alpha, \upsilon) \times \mathcal{M}(\alpha, \upsilon)$ at fixed α.

Finally,

$$\frac{1}{2}\left[\min\left\{1, \frac{1-\alpha}{\alpha}\right\} + \min\left\{1, \frac{\alpha}{1-\alpha}\right\}\right] = \begin{cases} \dfrac{1}{2(1-\alpha)}, & \alpha \le \dfrac{1}{2}, \\ \dfrac{1}{2\alpha}, & \alpha \ge \dfrac{1}{2}. \end{cases}$$

This expression is increasing on $(0, \frac{1}{2}]$, decreasing on $[\frac{1}{2}, 1)$, and reaches its maximum value 1 at $\alpha = \frac{1}{2}$. Since μ^* is continuous, $\upsilon(\alpha, \mu^*)$ is atomless for every $\alpha \in (0, 1)$, so the argument applies throughout the interior. □

References

Galichon, Alfred. 2018. *Optimal transport methods in economics.* Princeton University Press.

Horowitz, Joel L, and Charles F Manski. 1995. "Identification and robustness with contaminated and corrupted data." *Econometrica*, 281–302.

Lieb, Elliott H, and Michael Loss. 2001. *Analysis.* Vol. 14, American Mathematical Soc.

Manski, Charles F. 2004. "Social learning from private experiences: the dynamics of the selection problem." *The Review of Economic Studies*, 71(2): 443–458.