Fathers and Kings:
Who Gains from the Family Metaphor?

Niels Boissonnet*

Second draft, revised 12 November 2026

Abstract

Rulers have long called themselves fathers of their people. Who gains from the metaphor, and what does it do to families? In our model people judge husbands and rulers by one expectation about those in charge, right on average and wrong case by case, as in analogy-based expectation equilibrium. Any such analogy transfers trust from the authorities who behave better than their class to those who behave worse, and creates none. A ruler worse than fathers therefore gains from the family metaphor, and gains most where patriarchy is common, since fathers are then most of the authority people see. A ruler who chooses his analogy, which authority to resemble and how much, pools himself only with authorities better than he is, and they lose trust. Hence the metaphor he cultivates can discredit fathers but never launder them. It follows patriarchy where fathers deserve trust, and never sustains it where they do not. The metaphor sustains patriarchy in two cases only: when the ruler values patriarchy for itself, or when a tradition he cannot erase carries the analogy. In both, fathers do not deserve trust on their own, but the order around them does.

Keywords: analogy; authority; family structure; political trust; analogy-based expectations; patriarchy; propaganda.
JEL: D02, D83, D91, J12, Z13.

1 Introduction

In 1680 Robert Filmer argued that kings rule by the right of fathers: political authority is paternal authority writ large (Filmer, 1680). Locke answered that the two have nothing in common (Locke, 1689). The quarrel was about an analogy, and it turned on one thing: whether the one who holds power can be trusted with it. The analogy has outlived the quarrel. Tsars were little fathers, dictators have been fathers of the nation, and political language still casts the good ruler as a father and the tyrant as an abusive one. Two facts about this language stand out. Rulers use it; and rulers of patriarchal societies use it more than others.

We study who gains from the metaphor, and what does it do to families? Our premise is that people do not keep separate books for the different authorities in their lives. A husband who decides for his household and a ruler who decides for his country are both "the one in charge", and from either people expect the conduct they have learned to expect from those in charge in

*Center for Mathematical Economics, Bielefeld University. Email: niels.boissonnet@gmail.com.

general. We model this idea with the analogy-based expectation equilibrium of Jehiel (2005). People pool situations into classes—"holders of authority"– and expect, in every situation of a class, the average behavior over the class, weighted by how often each situation occurs.

The model has two kinds of decision. New couples, or those who decide for them, choose between a patriarchal marriage, in which the husband decides, and an egalitarian one, in which decisions are shared; they choose on what they expect of a husband's authority. Citizens grant the ruler more or less authority on what they expect of his. Under the analogy both expectations are one number: the trust that authority in general deserves, given the conduct of husbands, of the ruler and of other authorities, each weighted by how often people meet them. Conduct is a primitive. The weights move with society, since fewer patriarchal households show fewer husbands in charge and more participation shows less of the ruler. The metaphor sets the rest: which authorities the ruler is pooled with, and how much the pooling weighs.

Two results organize the paper.

The analogy is a transfer of trust. Any analogy that is right on average over encounters moves trust from the authorities who behave better than their class to those who behave worse, and creates none (Lemma 1). A ruler therefore gains from a metaphor exactly when it pools him with authorities better than himself, and he gains the more, the better they behave and the more often they are met (Proposition 1). The family metaphor is worth most to a ruler worse than fathers, in a society where fathers deserve trust and patriarchy is common, because fathers are then most of the authority people see. Benevolent fathers pay for it.

The metaphor power cultivates never launders. A ruler who chooses his analogy to gain authority pools only with authorities better than the class he forms, so whoever he borrows from loses trust. He can discredit fathers through the metaphor; he can never make them look better than they are (Theorem 1). Hence the metaphor he cultivates never sustains patriarchy against fathers' own conduct. Family structure ends where fathers' conduct sends it, and the metaphor follows patriarchy rather than creating it. The ruler maintains it where fathers deserve trust, so that families would be patriarchal anyway, and withdraws it where it would change families, because there it spends his reputation on others. In our draws, rulers cultivate the family analogy about six times more often in societies that end patriarchal than in societies that end egalitarian, which is the second fact we started from. The metaphor sustains patriarchy in two cases only: when the ruler values patriarchy for itself, or when a tradition he cannot erase carries the analogy. In both cases it does so only where fathers do not deserve trust on their own but the order around them, a benevolent state or a trusted church, does.

Both results rest on what the shared expectation does when it is taken as given. The ruler's conduct spills over onto families and husbands' onto politics, each in proportion to its share of the encounters with authority. A configuration, a society in which every household has the same form and participation has settled, hides the evidence against it: a patriarchal society sees benevolent husbands every day, and an egalitarian one sees mostly the state (Proposition 2). When husbands deserve trust, a patriarchal society with a strong ruler and an egalitarian one with participation are both self-consistent, and history decides between them (Proposition 3). This is one account of the correspondence between family structure and political structure that

Todd (1985) documents. Without the analogy nothing links the two spheres. The analogy can also keep a family form alive against husbands' own conduct, by lending them someone else's reputation (Proposition 5); the main result shows that a ruler who controls the analogy will not let it do so for patriarchy.

The analogy-based expectation is the steady state of learning when people record outcomes by class, not by holder (Jehiel, 2022): the Berk–Nash equilibrium of a learner who takes all holders of authority to share one type (Esponda and Pouzo, 2016). Its signature is a spillover across spheres that does not fade with experience, which a correctly specified learner would let fade (Proposition 7). Section 7 proposes tests of the discourse and of the spillover.

Related literature. The transfer of trust is the "transference" of Mullainathan et al. (2008), where coarse thinkers carry the content of a message across the situations of a category, and the ruler's choice of the audience's category is their "framing". Ettinger and Jehiel (2010) have a sophisticated player deceive coarse opponents, and Jehiel and Samuelson (2012) have analogical short-run players judge a long-run player by his average action; a sophisticated ruler facing analogical citizens is admissible in Jehiel's concept, since players may hold different partitions. Schwartzstein and Sunderam (2021) have persuaders propose models that fit the data. A consistent analogy fits the aggregate record exactly, which is why it is available to a persuader and why aggregate evidence does not dislodge it. What we add is that the class the ruler joins is made of people whose behavior responds to the trust placed in them, so the metaphor changes the composition of its own class. This feedback is what makes the metaphor follow patriarchy, and what produces cycles. The learning foundation of analogy-based expectations is in Jehiel (2022) and, experimentally, in Huck et al. (2011); Steiner and Stewart (2008) study similarity-weighted learning across games, and Mohlin (2014) the categorization that minimizes prediction error. Guriev and Treisman (2020) model rulers who survive by convincing the public of their competence; our ruler borrows trust rather than manufacturing signals. On families and politics, Todd (1985) documents the correspondence, Alesina and Giuliano (2011) the link between strong family ties and low participation, and Bisin and Verdier (2001, 2024) model the joint evolution of culture and institutions through socialization. Our link is a shared expectation, and it predicts that the state's conduct moves families with no policy at all. Tversky (1977) documents the asymmetry of similarity judgments, which appears here as an outcome of exposure. Linguistics and political psychology have described the political uses of the family metaphor (Lakoff, 1996); we offer a theory of who uses it, where, and what it does.

Section 2 presents the model; Sections 3 and 4 take the analogy as given, Sections 5 and 6 let the ruler choose it, Section 7 discusses tests, and the Appendix contains the dynamics of the joint system, the proofs and the numerical methods.

2 The model

Society. The state of society is a pair (β, x). The share of egalitarian households is $x \in [0,1]$; the rest are patriarchal. Political participation is $\beta \in [b_\ell, b_h] \subset (0,1)$, and the ruler holds authority $1-\beta$. The bounds mean that institutions never give the ruler all power, nor citizens all of it.

Holders of authority. Three kinds of people exercise authority over others: the husband in a patriarchal household (F), the ruler (S), and other authorities such as employers, clergy or elders (O). An egalitarian household has no holder: decisions are shared, so no one exercises authority over another. A holder of kind k is *benevolent* with probability π_k, meaning that he uses his authority in the interest of those subject to it; otherwise he abuses it. Conduct is a primitive, identical within each kind. Appendix A lets the ruler choose his.

Exposure. People learn about authority from the authority they experience or witness: the households they grew up in and around, the state's interventions, and other authorities. We measure exposure to each kind by

$$f_F = \phi_F(1-x), \qquad f_S = \phi_S(1-\beta), \qquad f_O = \nu, \qquad \Sigma = f_F + f_S + f_O. \tag{1}$$

Authority at home is met in proportion to the share of patriarchal households. State authority is met in proportion to the authority the ruler holds, since a ruler with more power intervenes in more lives. Exposure to other authorities is constant. The intensities ϕ_F and ϕ_S scale how often each kind is met. We have in mind $\phi_F > \phi_S$, since authority at home is daily and the state's occasional, but nothing below requires it.

Analogies. An analogy pools kinds of holders into classes. People expect every holder in a class to be benevolent with the class average probability, weighted by exposure. This is the consistency condition of analogy-based expectation equilibrium (Jehiel, 2005): if one counted all encounters with the holders of a class and the share of benevolent ones, one would find the class average. The error lies in applying the average to each holder. Two analogies bound the others. Under the *fine partition* each kind is its own class and expectations are correct holder by holder, $e_F = \pi_F$ and $e_S = \pi_S$; we call it the benchmark. Under the *coarse partition* all holders are in one class and

$$e_F = e_S = \hat{\pi} = \frac{f_F\pi_F + f_S\pi_S + f_O\pi_O}{\Sigma}. \tag{2}$$

Between the two, let a share $\omega \in [0,1]$ of the population pool all holders in one class while the rest keep separate books. The average expectation about holders of kind k is then

$$e_k = (1-\omega)\,\pi_k + \omega\,\hat{\pi}, \tag{3}$$

and we call ω the *weight of the analogy*. Jehiel's concept allows players to hold different partitions, so this is an analogy-based expectation equilibrium with two kinds of citizens. Sections 3 and 4 take the analogy as given, at full weight unless said otherwise. Section 5 lets the ruler choose which kinds he is pooled with, and the weight.

The third kind of holder is not decorative. Without it, a society with only egalitarian households and full participation would show no authority at all, and expectations would be undefined. With it, expectations are always defined and correct on average. A prior in its place would make them wrong on average, which Jehiel's concept excludes.

Families. A couple that forms chooses the form of its marriage. We treat the bride's consent as decisive, since she is the one subject to a husband's authority.[1] In a patriarchal marriage the husband decides. If he is benevolent, single-headed decisions bring a gain g, from speed and specialization; if he abuses them, a loss ℓ. In an egalitarian marriage decisions are shared and yield u_E. With expected benevolence e_F, the egalitarian form is better by

$$u_E - \big[e_F\, g - (1 - e_F)\, \ell\big] = k_F\,(\pi^* - e_F), \qquad k_F = g + \ell, \quad \pi^* = \frac{u_E + \ell}{g + \ell}.$$

The payoff gap is linear in trust, and π^* is the *trust threshold*: the level of trust at which the two forms are equally good. Households form and dissolve at rate λ, and existing households keep their form. A new couple picks the egalitarian form with the logit probability $\Lambda(\kappa k_F(\pi^* - e_F))$, where $\Lambda(z) = 1/(1 + e^{-z})$ and κ measures how sharply choices respond to expected payoffs. Hence

$$\dot{x} = \lambda\Big[\Lambda\big(\kappa k_F(\pi^* - e_F)\big) - x\Big]. \tag{4}$$

The choice must rest on expected payoffs. If forms spread by imitating those that succeed, success would be measured by realized payoffs, which depend on π_F, and expectations would play no role.

We use the average expectation (3) in both decisions. This is exact for participation, whose rule below is linear, and an approximation for new couples, whose choice is logistic. The approximation costs little, and the results of Section 5 are the same when each kind of citizen chooses on her own expectation. Over 20,000 random draws with the ruler setting the weight, the end state follows the sign of $\pi_F - \pi^*$ in 99.5 to 99.8% of draws under both readings. The population reading ends in a mixed society in 2 to 3% of draws, against under 1% under the average expectation (Appendix C).

Politics. Granting the ruler authority pays when he uses it well and costs when he abuses it, so citizens want less participation when they trust him more. We take the simplest such rule, linear and bounded,

$$b(e) = \text{clip}\,\big(\beta_m - s\,(e - \pi^*),\, b_\ell,\, b_h\big), \qquad \dot{\beta} = \eta\,\big[b(e_S) - \beta\big], \tag{5}$$

where s measures how much participation responds to trust and $\beta_m = b(\pi^*)$ is the participation chosen when trust is at the threshold. Participation adjusts gradually toward its target. The target is a reduced form for what the political process delivers, not an individual choice. Using the family's trust threshold in politics is a simplification; a different threshold, or a convex or concave rule, changes no result (Appendix C).

[1] To speak of the bride's choice is a simplification, and an uncomfortable one. For most of history women had little say in the form of their marriage, and often none in the choice of a husband: parents, kin, custom and law decided for them. The model does not need the bride to be free. It needs only that patriarchal marriages become rarer when husbands' authority is trusted less, whoever decides. The decisive party may be the bride's parents, who weigh the risk that a son-in-law abuses his authority over their daughter, or a custom that grants or withholds a husband's rights. Constraints on the choice lower its sharpness κ; sanctions against the egalitarian form lower its payoff u_E, hence the threshold π^* defined below. Neither changes the logic. What the model excludes is a decider whose payoff does not depend on how husbands use their authority. Nor does it claim that a woman who enters a patriarchal marriage endorses it: it may be the best of poor options.

Time. Time is continuous. Families and politics adjust at the same time, each at its own speed. The unit is set by political adjustment ($\eta = 1$); a household generation lasts $1/\lambda$. Along a path conduct is fixed, so only exposure moves expectations: f_F falls as egalitarian households spread, and f_S falls as participation rises. This is the whole dynamic of the model, until the ruler acts on the analogy.

Baseline. For the numerical illustrations we use $\pi_F = 0.8$, $\pi_S = 0.5$, $\pi_O = 0.3$; $\phi_F = 1$, $\phi_S = 0.5$, $\nu = 0.5$; $\pi^* = 0.55$, $k_F = 2$, $\kappa = 50$, $\lambda = 0.5$; $\beta_m = 0.325$, $s = 1.5$, $[b_\ell, b_h] = [0.15, 0.5]$. Husbands are mostly benevolent, the ruler middling, other authorities poor. The parameters are illustrative; the Appendix reports sweeps over wide ranges.

What the model leaves out. It is silent on why authorities behave as they do, except for the ruler. It has one representative exposure: men and women, who stand on different sides of a husband's authority, hold the same expectation. It has no family policy: the state reaches families through expectations only. This is deliberate. It isolates the channel we study, and a policy channel can be added.

3 Who gains from an analogy

Rulers of patriarchal societies have often presented themselves as fathers of their people. Why would the patriarchal side have more interest in the analogy than the other? The answer needs no dynamics, only exposures and conduct, and we give it first.

3.1 A transfer of trust

Write an analogy as a matrix W of nonnegative weights with rows summing to one, so that $e_k = \sum_j w_{kj}\pi_j$. Call it *consistent* if expectations are right on average over encounters, $\sum_k f_k e_k = \sum_k f_k \pi_k$, whatever the conducts. This is Jehiel's condition, required here for every conduct vector rather than at the realized one. Every partition and every mixture of partitions satisfies it, so (3) does at every ω.

Lemma 1 (A transfer of trust)**.** *An analogy is consistent if and only if $fW = f$. Then, for every conduct vector, $\sum_k f_k (e_k - \pi_k) = 0$: weighted by exposure, the gains and losses of trust sum to zero. Under the coarse partition, moving from $\omega = 0$ to $\omega = 1$ changes the trust placed in holders of kind k by $G_k = \hat{\pi} - \pi_k$, so a holder gains if and only if his conduct is below the class average.*

An analogy creates no trust. It moves trust from the holders who behave better than their class to those who behave worse, and it moves nothing else. A similarity-weighted expectation, in which the ruler resembles fathers to some degree and expectations are kernel averages, is not consistent in general (Steiner and Stewart, 2008): it can create trust out of nothing. This is why we stay with partitions and their mixtures. The baseline gives the orders of magnitude. Take its patriarchal configuration, the state of Section 4 in which every household is patriarchal and participation has settled, where $\hat{\pi} = 0.606$. The coarse analogy gives the ruler 0.106 in trust

and, through the political rule, 0.159 in authority; other authorities gain 0.306; husbands lose 0.194. The beneficiaries are not fathers. Benevolent fathers pay for being pooled with the state.

3.2 Which metaphor

The coarse partition is one analogy among many. Let the ruler choose the set C of other kinds pooled with him, kinds outside C being judged on their own conduct. His class average is

$$e_S(C) = \frac{f_S\pi_S + \sum_{j\in C} f_j\pi_j}{f_S + F_C}, \qquad F_C = \sum_{j\in C} f_j,$$

and his gain in trust is $G_S(C) = e_S(C) - \pi_S$. Since authority is nonincreasing in trust, he wants the class with the highest average.

Proposition 1 (The choice of metaphor)**.** *(i) Adding a kind j to the class raises its average if and only if $\pi_j > e_S(C)$. A class maximizes the average if and only if it is a top set, $\{j : \pi_j > \lambda\} \subseteq C \subseteq \{j : \pi_j \geq \lambda\}$ with $\lambda = e_S(C)$; the top set is found by adding kinds in decreasing order of conduct while they exceed the running average. At it $e_S(C) \geq \pi_S$, and every pooled kind satisfies $\pi_j \geq e_S(C)$, strictly unless its conduct equals the average: whoever the ruler borrows from loses trust. If each metaphor carries a cost, the maximizer is no longer a top set, but every pooled kind still satisfies $\pi_j \geq e_S(C)$.*

(ii) The gain from the family metaphor alone, $G_S(\{F\}) = f_F\,(\pi_F - \pi_S)/(f_S + f_F)$, is increasing in π_F, and, when $\pi_F > \pi_S$, increasing in $f_F = \phi_F(1-x)$ and decreasing in f_S.

(iii) Between two metaphors taken alone, each pooling him with one kind better than he is, the ruler prefers the kind met more often when the two have equal conduct, and the better kind when the two have equal exposure.

Part (i) identifies whom the ruler pools with: the authorities that behave better than the class he forms, and only them. Part (ii) is the patriarchal side's interest. The family metaphor is worth more to a ruler the more patriarchal households there are, since more of the authority people see is then fathers', and the better fathers behave. In particular, when the two configurations of Section 4 are both stable, the ruler's gain from the coarse analogy is larger in the patriarchal configuration than in the egalitarian one: $G_S^P - G_S^E = \hat{\pi}(\beta_P, 0) - \hat{\pi}(\beta_E, 1) > 0$, where β_P and β_E are the levels of participation at the two configurations. Across 573 random draws with two stable configurations, the gain is larger in every draw, and the husbands of the patriarchal configuration lose in every draw. Part (iii) is a theory of which metaphor. A ruler uses every metaphor whose referent is better than his class. Among single metaphors he calls himself a father where fathers are the best-regarded and most-met authority, a shepherd where the clergy is, and nothing where he is himself the best authority people know. Over 20,000 draws in the ranges of the Appendix, at the patriarchal configuration and with the cost of propaganda of Section 5, the ruler uses no metaphor in 72% of draws, the family alone in 25%, other authorities alone in 2%, and both in 1%. He uses the family metaphor in 45% of the draws where fathers are better than he is, and in none where they are worse.

A ruler who also chooses his conduct gains twice. Under the analogy he bears only the share f_S/Σ of the trust he creates or destroys, the rest being credited to husbands, so he behaves worse than he would without it (Proposition 10 in the Appendix). In the baseline patriarchal configuration he chooses a benevolence of 0.25 under the analogy against 0.67 without it. The analogy gives him trust he did not earn, and lowers his cost of behaving badly.

4 What the analogy does when it is given

Before letting the ruler choose his analogy, we need to know what the shared expectation does to families and politics when it is taken as given, at full weight. Three facts matter for what follows. Conduct spills over from one sphere to the other; a configuration hides the evidence against it; and the analogy can keep a family form alive against husbands' own conduct, by lending them someone else's reputation. The dynamics of the joint system, which these facts do not need, are in Appendix A.

4.1 Spillovers and self-concealment

Write $A(\beta) = [f_S\pi_S + \nu\pi_O]/(f_S + \nu)$ for the average conduct of the authority met outside the family.

Proposition 2 (Spillovers and self-concealment)**.** *Under the coarse partition,*

$$\frac{\partial e_F}{\partial \pi_S} = \frac{f_S}{\Sigma} > 0, \qquad \frac{\partial e_S}{\partial \pi_F} = \frac{f_F}{\Sigma} > 0, \qquad \frac{\partial e_S/\partial \pi_F}{\partial e_F/\partial \pi_S} = \frac{\phi_F(1-x)}{\phi_S(1-\beta)},$$

and $\partial\hat{\pi}/\partial x = -\phi_F(\pi_F - \hat{\pi})/\Sigma$, $\partial\hat{\pi}/\partial\beta = -\phi_S(\pi_S - \hat{\pi})/\Sigma$. *For every* β, $\hat{\pi}(\beta, \cdot)$ *is monotone in* x*: decreasing if* $\pi_F > A(\beta)$*, increasing if* $\pi_F < A(\beta)$*. Likewise* $\hat{\pi}(\cdot, x)$ *is monotone in* β*. Under the benchmark the cross effects are zero and* $\hat{\pi}$ *is replaced by constants.*

A more benevolent ruler raises the trust placed in husbands, and more benevolent husbands raise the trust placed in the ruler, each in proportion to its share of the encounters with authority. Expectations respond more to the family than to the state as long as authority is met more often at home. In the baseline the ratio is 2.5 in a patriarchal society ($\beta = 0.2$, $x = 0$), 1.4 in a mixed one, and 0.18 in an egalitarian one ($\beta = 0.45$, $x = 0.95$). The asymmetry that Tversky (1977) documents for similarity judgments, where the variant is judged by the prototype more than the reverse, appears here as an outcome of exposure, and it reverses once patriarchy is rare. The second part is self-concealment. When husbands are better than the authority met outside the family, each egalitarian household removes a good example of authority from view, and trust falls: egalitarian marriage gains from its own spread, through expectations alone. Symmetrically, a patriarchal society sees benevolent husbands every day and finds authority trustworthy. The direction never changes along the way, because fathers are compared with a fixed outside average.

4.2 Configurations

Call $(\beta_P, 0)$ with $\beta_P = b(\hat{\pi}(\beta_P, 0))$ the *patriarchal configuration* and $(\beta_E, 1)$ with $\beta_E = b(\hat{\pi}(\beta_E, 1))$ the *egalitarian configuration*, and write $\hat{\pi}_P = \hat{\pi}(\beta_P, 0)$, $\hat{\pi}_E = \hat{\pi}(\beta_E, 1)$. In the

sharp-choice limit $\kappa \to \infty$ new couples choose equality if and only if $e_F < \pi^*$. A configuration is *self-confirming* when new couples at it choose its form, and *stable* when in addition participation settles there.

Proposition 3 (Configurations). *(i) For every x, participation has a steady state $\beta = b(\hat{\pi}(\beta, x))$ in $[b_\ell, b_h]$. If*

$$s\,\phi_S\,\frac{M-m}{\phi_F(1-x)+\phi_S(1-b_h)+\nu} < 1, \tag{6}$$

where $M = \max_k \pi_k$ and $m = \min_k \pi_k$ are the best and the worst conduct, it is unique and participation converges to it from every start. In particular both configurations exist.

(ii) For given β and finite κ: if $\pi_F > A(\beta)$, the share of new couples who choose equality rises with the share that already have, and the family dynamics has two stable steady states, one near each corner, as soon as κ exceeds an explicit threshold (Proposition 9 in the Appendix); if $\pi_F \le A(\beta)$, it has a unique steady state, which attracts every path.

(iii) In the sharp-choice limit, under (6), the patriarchal configuration is stable if and only if $\hat{\pi}_P > \pi^*$ and the egalitarian one if and only if $\hat{\pi}_E < \pi^*$. Both are stable if and only if

$$\hat{\pi}_P \;>\; \pi^* \;>\; \hat{\pi}_E. \tag{7}$$

Then $\beta_E \ge \beta_P$, and fathers deserve trust, $\pi_F > \pi^*$, unless the ruler is better than other authorities and participation is strictly higher in the egalitarian configuration. Under the benchmark the two never coexist: the share goes to 0 if $\pi_F > \pi^*$ and to 1 if $\pi_F < \pi^*$, from every start.

Part (ii) is the dividing line of the model: are fathers better or worse than the authority people meet outside the family? If they are better, trust falls as egalitarian households spread, so the choice of equality feeds on itself and two family configurations can coexist. If they are worse, their disappearance raises trust, the choice is self-limiting, and the outcome is unique. Part (iii) gives what coexistence requires. A patriarchal society trusts authority enough to stay patriarchal, because the authority it sees, mostly fathers, deserves trust; an egalitarian society distrusts authority enough to stay egalitarian, because the authority it sees, the ruler and other authorities, does not. Each configuration is self-consistent because it hides the evidence against it. The exception, two configurations through politics alone, needs a strong response of participation to a ruler better than other authorities. An example exists, but none of 420,000 random draws over our ranges shows it (Appendix C), and we set it aside.

Which configurations exist can be read off two numbers:

$$\Delta_F = \phi_F\,(\pi_F - \pi^*), \qquad \Delta_O = \phi_S(1-\beta_m)(\pi^* - \pi_S) + \nu\,(\pi^* - \pi_O).$$

The first is the trust that fathers produce beyond the threshold when every household is patriarchal. The second is the trust missing outside the family, from the ruler and from other authorities, at the participation β_m that citizens choose when trust is at the threshold.

Proposition 4 (Configurations in primitives)**.** *Under* (6)*, participation at a configuration is below* β_m *if and only if trust there is above* π^**, and if and only if* $\hat{\pi}(\beta_m, \cdot)$ *is above* π^* *at the same corner. Hence, in the sharp-choice limit, the patriarchal configuration is stable if and only if* $\Delta_O < \Delta_F$*, and the egalitarian one if and only if* $\Delta_O > 0$*. Both are stable if and only if* $0 < \Delta_O < \Delta_F$*; neither is if and only if* $\Delta_F < \Delta_O < 0$*. In the last case society settles at a mixture: as* $\kappa \to \infty$ *every steady state tends to the state with participation* β_m *and share* $1 - \Delta_O/\Delta_F$*, and is locally stable for* κ *large.*

A society that trusts authority grants it more. So participation at a configuration is below β_m exactly when trust there is above π^*, and trust can be compared with the threshold at β_m, without computing any fixed point. Two configurations coexist when trust is missing outside the family, but less of it than fathers produce. Neither the response of participation to trust, nor the speeds of adjustment, nor the bounds on participation enter the two numbers: politics decides nothing about which configurations exist.

This is one account of why the state resembles the family. The baseline with $\kappa = 50$ has three steady states: a stable patriarchal one at $(\beta, x) = (0.241, 0.004)$ with $\hat{\pi} = 0.606$, a stable egalitarian one at $(0.500, 1.000)$ with $\hat{\pi} = 0.367$, and a saddle between them. Under the benchmark every start ends at $(0.40, 0)$: families go where husbands' conduct sends them, politics where the ruler's sends it, and nothing links the two. Under the analogy, history decides between a strong ruler with patriarchal households and participation with egalitarian ones, which is the correspondence Todd (1985) documents. Over 20,000 random parameter vectors drawn over wide ranges, the two configurations coexist in about three draws in ten at the draws' own sharpness of choices. Proposition 4 is a limit statement and misses one requirement: at finite sharpness, coexistence also needs choices sharp enough (Proposition 9 in the Appendix). Learning by exposure without the analogy, which keeps separate classes but learns from what is seen, also produces self-concealment and coexistence; what it lacks is the link between the spheres, so that participation is the same in both configurations.

4.3 Four cases

Proposition 3(iii) describes configurations in which fathers deserve trust on their own. The analogy can also work the other way: fathers who do not deserve trust can borrow it from a better state. Comparing the configurations reached with the analogy and without it, four things can happen.

Proposition 5 (Four cases)**.** *Under* (6)*, in the sharp-choice limit:*

(i) an egalitarian configuration exists with the analogy and not without it iff $\pi_F > \pi^* > \hat{\pi}_E$*;*

(ii) a patriarchal configuration exists without the analogy and not with it iff $\pi_F > \pi^* > \hat{\pi}_P$*;*

(iii) an egalitarian configuration exists without the analogy and not with it iff $\pi_F < \pi^* < \hat{\pi}_E$*;*

(iv) a patriarchal configuration exists with the analogy and not without it iff $\pi_F < \pi^* < \hat{\pi}_P$*.*

In cases (iii) and (iv) the configuration concerned is the unique family steady state at its participation; in case (i) it can coexist with the patriarchal one.

Cases (i) and (iv) are the analogy *sustaining* a family form against husbands' own conduct: equality against trusted fathers, judged by a worse state; patriarchy against mediocre fathers, who borrow the trust of a better state. In the terms of Proposition 4, case (i) is $\Delta_F > 0$ and $\Delta_O > 0$, case (ii) is $\Delta_O > \Delta_F > 0$, case (iii) is $\Delta_F < 0$ and $\Delta_O < 0$, and case (iv) is $\Delta_O < \Delta_F < 0$. The two sustaining cases are not symmetric. Case (iv) requires $-\Delta_O > -\Delta_F$, that is $\phi_F(\pi^* - \pi_F) < \phi_S(1-\beta_m)(\pi_S - \pi^*) + \nu(\pi_O - \pi^*)$: the fathers' deficit, weighted by the whole family's exposure, must be outweighed by the trust produced outside the family. Case (i) requires only $\Delta_O > 0$ besides trusted fathers, with no family term, since in an egalitarian society no father exercises authority. The analogy transmits distrust fully once fathers are out of sight, and launders mediocre fathers only partly while they are everywhere. Over 50,000 draws covering both rankings of conduct, the analogy creates equality in 37% of draws and patriarchy in 2%; in a world where the state and other authorities are benevolent and fathers mediocre, it creates patriarchy in 36% (Appendix C). Which world a society is in is an empirical question about the ranking of conduct. Whether a ruler would let the analogy sustain patriarchy is the question of the next section.

5 The metaphor power cultivates never launders

Rulers do not take the analogy as given. They call themselves fathers, or shepherds, or they insist on being judged on their record. We now let the ruler choose his analogy along the two margins of Section 3: which kinds he is pooled with, and how much weight the pooling carries. Both cost him something, which we call propaganda. He maximizes the authority citizens grant him, net of this cost, taking the current state of families and participation as given. He is thus myopic about families; we relax this below.

5.1 The main result

Write $e_F = (1-\omega)\pi_F + \omega\, e_S(C)$ for the expectation about husbands when $F \in C$ and the pooling has weight ω, and $e_F = \pi_F$ otherwise.

Theorem 1 (The metaphor never launders)**.** *Let the ruler choose the class $C \subseteq \{F, O\}$ and the weight $\omega \in [0,1]$ to maximize*

$$1 - b\big((1-\omega)\pi_S + \omega\, e_S(C)\big) - \tfrac{c_\omega}{2}\,\omega^2 - \sum_{j \in C} \kappa_j, \qquad c_\omega > 0,\ \kappa_j \geq 0,$$

a cost of intensity that does not depend on the class and a cost per metaphor, kinds outside C being judged on their own conduct. Take the sharp-choice limit.

(i) At any optimum with $\omega > 0$, the class average exceeds the ruler's conduct, $e_S(C) > \pi_S$, and every pooled kind satisfies $\pi_j \geq e_S(C)$. Hence whenever fathers are pooled, $e_F \leq \pi_F$, whatever the weight and whether or not participation is at a bound.

(ii) If $\pi_F < \pi^*$, no patriarchal configuration is self-confirming: new couples choose equality whatever the ruler does and whatever the other authorities do.

(iii) If $\pi_F > \pi^*$, the patriarchal configuration is self-confirming if and only if $e_F > \pi^*$, which fails only when fathers are pooled and the class discredits them below the threshold. The egalitarian configuration is self-confirming only if fathers are pooled at a positive weight and the class contains another kind better than the ruler, $\pi_O > \pi_S$; since no father exercises authority there, fathers add nothing to the class average, and the ruler keeps them in it only if $\kappa_F = 0$.

The proof is short. A ruler who uses a metaphor at all does so because his class is better than he is. A pooled kind worse than the class could be dropped, and the same expectation about the ruler reached at a lower weight and a lower cost. So every kind he borrows from is at least as good as the class, and is judged by the class: it loses trust. Fathers in particular. An authority-seeking ruler can discredit fathers through the metaphor; he can never make them look better than they are. The given analogy of Section 4 could sustain patriarchy against mediocre fathers, case (iv) of Proposition 5, by lending them the trust of a better state. The analogy a ruler cultivates cannot. A ruler who wanted to borrow from better authorities would not take mediocre fathers into his class, since they dilute it at every state of society. Part (iii) locates the asymmetry. Trusted fathers are worth pooling with, and pooling can discredit them below the threshold; so the metaphor can destroy patriarchy, and it can sustain equality against trusted fathers when the ruler also borrows from better authorities than himself. This is case (i) of Proposition 5, and it is where the cycles below live. At the egalitarian corner, with fathers kept in the class whenever it is free, equality is self-confirming against trusted fathers in 0.7% of our draws, all with $\pi_O > \pi_S$ (Appendix C). Fathers worse than the class the ruler would form are never pooled, and pooled fathers are always judged below their conduct; so the metaphor never sustains patriarchy. The sentence that summarizes the theorem is the title of this section.

Two assumptions carry the result. The cost of intensity does not depend on the metaphor: if calling oneself a father were much cheaper than calling oneself a shepherd, a ruler could find it worth pooling mediocre fathers together with a better clergy, and launder them in passing. And the ruler controls only his own class: he cannot pool fathers with better authorities without joining the class himself. Within these assumptions, the result holds for any cost of the metaphors, at any state of society, and with participation at a bound or not.

5.2 The weight of the analogy, with the class given

The simulations below use the technology of (3): the class is all holders of authority, and the ruler sets its weight. At each moment he picks the weight that maximizes his authority net of a quadratic cost, and the weight adjusts toward it:

$$\omega^*(\hat{\pi}) = \arg\max_{\omega\in[0,1]} \Big[1 - b\big(e_S(\omega)\big) - \tfrac{c_\omega}{2}\,\omega^2\Big], \qquad \dot{\omega} = \zeta\,\big(\omega^*(\hat{\pi}) - \omega\big), \tag{8}$$

with $e_S(\omega) = (1-\omega)\pi_S + \omega\hat{\pi}$.

Lemma 2 (The ruler's use of the analogy)**.** *Let $\bar{e} = \pi^* + (\beta_m - b_\ell)/s$ be the trust at which participation reaches its floor. Whatever the bounds, the ruler does not use the analogy when his conduct is at least the class average, and otherwise $\omega^* \le \min\{1,\, s(\hat{\pi} - \pi_S)/c_\omega\}$. If participation*

is not at its ceiling when he is judged on his own conduct, $\beta_m + s(\pi^* - \pi_S) \leq b_h$, his problem is concave and, for $\hat{\pi} > \pi_S$,

$$\omega^*(\hat{\pi}) = \min\Big\{1,\ \frac{s\,(\hat{\pi} - \pi_S)}{c_\omega},\ \frac{(\bar{e} - \pi_S)^+}{\hat{\pi} - \pi_S}\Big\}.$$

The ruler uses the analogy when the class average exceeds his own conduct, the more so the cheaper propaganda is, and no further than the weight at which he holds all the authority the institutions allow. We say that there is *no ceiling* when the condition of the lemma holds, as in the baseline. It fails for a ruler so far below the threshold that his own conduct puts participation at its upper bound. The analogy then pays only once it has earned the trust he lacks, his best weight jumps from zero, and a configuration can fail to exist (Appendix B).

The joint system is now three-dimensional, in (β, x, ω). Its configurations can be read off exactly in the sharp-choice limit.

Theorem 2 (Families follow fathers)**.** *Let $\kappa \to \infty$, $\pi_F \neq \pi^*$, and let there be no ceiling. A configuration under the ruler's choice is a point (β, x, ω) with $x \in \{0, 1\}$, $\omega = \omega^*(\hat{\pi})$ and $\beta = b((1-\omega)\pi_S + \omega\,\hat{\pi})$, where $\hat{\pi} = \hat{\pi}(\beta, x)$. Write $A_P = (\phi_F \pi_F + \nu \pi_O)/(\phi_F + \nu)$ and $A_E = \pi_O$ for the average conduct of the holders other than the ruler at each corner, and $\hat{\pi}_P$, $\hat{\pi}_E$ for the class average at each configuration.*

(i) (One configuration.) There is exactly one patriarchal and one egalitarian configuration. At a fixed corner, participation and the weight of the analogy converge to it from every start.

(ii) (Neutrality.) The ruler uses the analogy at the patriarchal configuration if and only if $\pi_S < \min\{A_P, \bar{e}\}$, and at the egalitarian one if and only if $\pi_S < \min\{\pi_O, \bar{e}\}$. Where he does not, the configuration is self-confirming if and only if fathers' own conduct makes it so. If $\pi_S \geq \max\{A_P, \pi_O\}$ he uses it at no state of society, and every path ends at the corner that fathers' conduct supports.

(iii) (Trusted fathers, $\pi_F > \pi^*$.) The patriarchal configuration is self-confirming unless $\hat{\pi}_P \in I_D$, and the egalitarian one is self-confirming if and only if $\hat{\pi}_E \in I_D$, where

$$I_D = \big\{p < \pi^* :\ s\,(p - \pi_S)(\pi_F - p) > c_\omega\,(\pi_F - \pi^*)\big\}.$$

I_D is an open interval or empty; it is empty if $\pi_S \geq \pi^*$ or if $c_\omega \geq s\,(\pi_F - \pi_S)^2/[4(\pi_F - \pi^*)]$.

(iv) (Mediocre fathers, $\pi_F < \pi^*$.) The egalitarian configuration is self-confirming unless $\hat{\pi}_E \in I_L$, and the patriarchal one is self-confirming if and only if $\hat{\pi}_P \in I_L$, where

$$I_L = \big\{p > \pi^* :\ s\,(p - \pi_S)(p - \pi_F) > c_\omega\,(\pi^* - \pi_F)\ \ \textit{and}\ \ (\bar{e} - \pi_S)(p - \pi_F) > (\pi^* - \pi_F)(p - \pi_S)\big\}.$$

I_L is an open interval or empty; it is empty if $\pi_S \geq \bar{e}$. A class average in I_L requires $\pi_O > \max\{\pi_S, \pi^*\}$ and $c_\omega < s\,(\pi_O - \pi_S)(\pi_O - \pi_F)/(\pi^* - \pi_F)$.

The theorem makes three points. First, the ruler's choice removes the political multiplicity of Section 4: a ruler better than the class average drops the analogy, is judged on his own

conduct, and participation is $b(\pi_S)$. Second, primitives decide who uses the analogy: the ruler who is worse than the other holders people meet at the configuration, and not already at the most trust he can use. When fathers are better than other authorities, $A_P > \pi_O$, so every ruler who uses the analogy in an egalitarian society also uses it in a patriarchal one, and those with $\pi_O \leq \pi_S < \min\{A_P, \bar{e}\}$ use it in the patriarchal one only. Third, families follow fathers except on two intervals of the class average. With trusted fathers, the analogy can undo patriarchy only if the class is below the threshold although fathers are above it, which needs a ruler below the threshold. It also needs propaganda cheap enough for the weight to reach the point where fathers are discredited. With mediocre fathers, it can sustain patriarchy only if other authorities are better than both the ruler and the threshold. The two intervals give a rule in primitives. Exactly one configuration is self-confirming, the one fathers' conduct supports, in three cases: fathers are trusted and either $\pi_S \geq \pi^*$ or $c_\omega \geq s(\pi_F - \pi_S)^2/[4(\pi_F - \pi^*)]$; fathers are mediocre and either $\pi_O \leq \max\{\pi_S, \pi^*\}$ or $c_\omega \geq s(\pi_O - \pi_S)(\pi_O - \pi_F)/(\pi^* - \pi_F)$; or $\pi_S \geq \max\{A_P, \pi_O\}$. Over the ranges of Appendix C these conditions hold in 98.7% of the draws without a ceiling.

Part (iv) is where the given class differs from the chosen one. With the class fixed to all holders, a ruler worse than other authorities can keep the analogy alive where it launders mediocre fathers, because he cannot borrow from the others without taking the fathers along. With the class chosen, he leaves the fathers out, and by Theorem 1(ii) this never happens. With trusted fathers the two technologies agree: when the ruler also chooses the class, the patriarchal configuration fails if and only if the average of the chosen class lies in I_D. The difference is small in numbers, since the case needs at once a ruler worse than other authorities, trusted authorities around mediocre fathers, and cheap propaganda. Over 20,000 draws at the patriarchal corner, it occurs in 2 of the 8,726 draws with mediocre fathers under the given class and in none under the chosen class. With trusted fathers the patriarchal configuration fails in 1.0% of draws under the chosen class and 1.4% under the given one (Appendix C). The simulations that follow therefore describe both technologies.

We also integrated the dynamics under the chosen class, the ruler picking his class and its weight at each instant, on 2,000 draws from both starts. With mediocre fathers, society ends patriarchal in no draw under either technology. With trusted fathers it ends with a majority of egalitarian households, the metaphor having discredited the fathers, in 1.3 to 2.1% of draws under the chosen class against 1.0 to 1.3% under the given one, and the two technologies give the same end in 99% of cases. The chosen class changes two things. First, it makes the cycles of propaganda described below rarer at the sharpness of our draws, without removing them. In the predatory-ruler baseline, the given class cycles because the other authorities drag the class average below the threshold; the chosen class leaves them out and settles at the patriarchal configuration with the family metaphor at full weight. But the situation that produces cycles, trusted fathers with neither configuration self-confirming, is almost as frequent under the chosen class (0.8% of the draws with trusted fathers and cheap propaganda, against 1.0% under the given class). With sharp choices society keeps moving there in about two draws in five; at the sharpness of our draws it settles at a mixture in most of them (Appendix C). Second, the chosen class makes the metaphor follow patriarchy more sharply: the ruler keeps an analogy in 42% of patriarchal ends, almost always the family alone, against 7% of egalitarian ends, where it is mostly the clergy (Appendix C).

regime	analogy at full weight ($\omega = 1$)	ruler sets ω
egalitarian only	57.8%	43.6%
patriarchal only	25.8%	55.9%
two stable configurations	16.4%	0.1%
cycles	—	0.4%

Table 1: Regimes over 10,000 draws covering both rankings of conduct, with the analogy fixed at full weight and with the ruler setting it.

5.3 Power cultivates the analogy where patriarchy already holds

Table 1 gives the outcome over 10,000 draws covering both rankings of conduct, with $c_\omega \in [0.02, 2]$ and $\zeta \in [0.05, 2]$. Two things happen when the ruler sets the analogy. The two configurations stop coexisting: among the 1,641 draws with two stable configurations, the manipulated society ends patriarchal in 99.5%. And the ending configuration follows fathers' own conduct: it is patriarchal if $\pi_F > \pi^*$ and egalitarian if $\pi_F < \pi^*$, in 98 to 100% of draws in each ranking of the ruler relative to fathers (Figure 1). Theorem 2 gives the exceptions; in the sharp-choice limit, over the same ranges, they make up 1.7% of the draws with trusted fathers and 0.3% of those with mediocre fathers. At the end the ruler's ω is zero in 78% of draws, interior in 20%, and one in 1.6%.

Power does maintain the analogy, however, where it rides on trusted fathers. Where the society ends patriarchal, the ruler keeps a positive weight on the analogy in 31% of draws; where it ends egalitarian, in 5%. With fathers who deserve trust and a ruler worse than them, the case of the paternal king, he keeps it in 40% of draws; the rest are draws where propaganda costs more than it brings, or where he already holds all the authority the institutions allow. With a ruler better than fathers, almost never. Once he can also choose his conduct, the picture is the same: he uses the analogy in 17% of patriarchal configurations and 1% of egalitarian ones, and where he uses it he behaves a little worse.

So the model reproduces the two facts we started from. Rulers cultivate the family analogy, and they do so in patriarchal societies, about six times more often than in egalitarian ones. Theorem 2(ii) gives the reason: whenever fathers are better than other authorities, the set of rulers who use the analogy at the patriarchal configuration contains the set of those who use it at the egalitarian one. In the sharp-choice limit the ruler uses it at 30% of the self-confirming patriarchal configurations and at 4% of the egalitarian ones. What the model adds is that the analogy an authority-seeking ruler cultivates *rides on* patriarchy rather than creating it. He maintains it where fathers deserve trust, so that families would be patriarchal anyway, and withdraws it where it would change families, because there it spends his reputation on others. The observable footprint is a correlation. With a ruler who sets the analogy, family structure follows fathers' own conduct; with an analogy that is given, it also follows the conduct of the state and of other authorities.

Cycles of propaganda. Theorem 2 leaves one situation open: neither configuration is self-confirming. With trusted fathers this happens when the class average at the patriarchal configuration lies in I_D and the one at the egalitarian configuration does not. The ruler pushes the

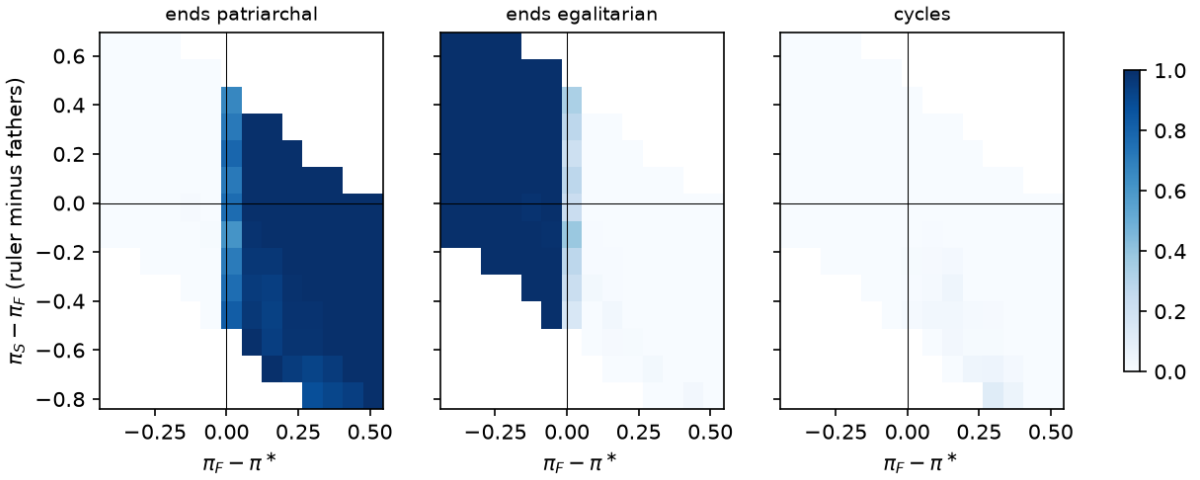


Figure 1: End configuration when the ruler sets the analogy, in the plane $(\pi_F - \pi^*,\ \pi_S - \pi_F)$ (cells with at least 10 draws). The frontier between the two outcomes is the line $\pi_F = \pi^*$, whatever the ruler's conduct, as Theorems 1 and 2(ii) predict.

analogy to borrow fathers' trust, which discredits them, and egalitarian households rise. With mediocre fathers it happens when the class average at the egalitarian configuration lies in I_L and the one at the patriarchal configuration does not. The two cases end differently.

Proposition 6 (Cycles and mixtures)**.** *Let there be no ceiling, and let neither configuration be self-confirming. As $\kappa \to \infty$, the steady states converge to one state, with both family forms present, at which husbands are judged exactly at the threshold and the class average is the lower end of I_D if $\pi_F > \pi^*$, of I_L if $\pi_F < \pi^*$.*

(i) If $\pi_F > \pi^*$, then for κ large enough no steady state is stable: society does not settle. At finite κ, a steady state at which the ruler uses the analogy and $\pi_F > \hat{\pi}$ is unstable if

$$\lambda\Big[\kappa k_F\, x(1-x)\,\frac{\phi_F(\pi_F - e_F)}{\Sigma} - 1\Big] > \zeta + \eta\Big[1 + \frac{s\,\phi_S\,(e_S - \pi_S)}{\Sigma}\Big]. \tag{9}$$

(ii) If $\pi_F < \pi^*$, then for κ large enough the steady state is locally stable: society settles at a mixture.

What decides is what the departing fathers take with them. Trusted fathers are better than the class the ruler pools them with. As patriarchal households disappear, the class average falls, husbands are trusted less, and more couples choose equality. The move feeds itself until the ruler, who adjusts with a delay, withdraws the analogy, and then it reverses. The metaphor discredits the fathers it borrows from, and the ruler's own target melts under him. Mediocre fathers are worse than the class. As patriarchal households come back, they lower the class average and the trust placed in husbands, which stops the move, and the interior state holds. Condition (9) identifies when cycles appear at finite sharpness: when households respond faster than the ruler's discourse and politics can follow. The simulations agree. In 0.45% of draws society cycles, all with fathers above the threshold, a ruler below it and cheap propaganda. The slower the ruler

adjusts his discourse, the longer and wider the cycles: at the lowest adjustment speed we tried, society swings almost fully between the two family forms over a period of about 62 time units. Cycles disappear when propaganda costs more than about a quarter of the authority at stake. They also need sharp choices: at the sharpness of our draws, most societies in the situation of part (i) settle at a mixture, and about one in five keeps moving (Appendix C). Under the chosen class none of our 2,000 draws cycles, though the situation of part (i) is almost as frequent and cycles reappear there when choices are sharp: fathers can also be discredited by a class that contains only them and the ruler. The ruler has a reason to avoid cycles. A ruler who commits once to a constant ω chosen to maximize long-run authority does the same as the myopic ruler in 95% of draws, within 10^{-3} of payoff, and where the myopic ruler cycles, commitment gains a median 0.049.

Occupiers. The same logic applies to an outside power that imposes a political structure for a time and governs with its own conduct: its conduct reaches families only through the analogy. In our draws, a predatory occupier switches a patriarchal society to egalitarian households in 46% of the draws with two stable configurations when it keeps the analogy at full weight, and in 0.6% when it drops it. Its abuses stay political unless people judge fathers by the authority it embodies.

6 When the cultivated analogy sustains patriarchy

Historically, the paternal metaphor seems to have done more than ride on patriarchy. By Theorem 1, an authority-seeking ruler cannot make it do more. We study two departures from him. An *ideological* ruler values patriarchy for itself: his payoff adds μ times the share of new couples who choose the patriarchal form, so that his target becomes $\arg\max_\omega[1 - b(e_S(\omega)) - \frac{c_\omega}{2}\omega^2 + \mu(1 - \Lambda(\kappa k_F(\pi^* - e_F(\omega))))]$. A *tradition* fixes a part τ of the analogy out of his reach, so that the weight that enters expectations is $\tau + (1-\tau)\omega$. In the sharp-choice limit the logic of Theorem 2 carries over with ω replaced by $\tau + (1-\tau)\omega^*$. Hence a tradition that puts husbands across the threshold on its own sustains the configuration whatever the ruler does: in case (iv), $\tau > (\pi^* - \pi_F)/(\hat{\pi}_P - \pi_F)$ with $\hat{\pi}_P > \pi^*$. An ideological ruler's objective has a cross derivative in (ω, μ) of the sign of $\hat{\pi} - \pi_F$, so he pushes the analogy beyond ω^* exactly where it makes husbands look better, which is the situation of case (iv). We ran 1,200 draws over both rankings of conduct for $\mu \in \{0, 0.2, 0.5, 1\}$ and $\tau \in \{0, 0.25, 0.5, 0.75\}$.

The analogy can sustain patriarchy at all only in case (iv): fathers below the threshold, trusted authority around them. It holds in 37 of the 1,200 draws, for the reasons of Section 4. In those draws, the share where the manipulated society ends patriarchal is

ideology μ	$\tau = 0$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
0	0.00	0.24	0.68	1.00
0.2	0.54	0.76	0.89	1.00
0.5	0.84	0.89	0.95	1.00
1	0.92	0.92	0.95	1.00

Without ideology or tradition, the analogy never sustains patriarchy, as Theorem 1 requires; with either, it mostly does; with a strong tradition, always. The sample is small, and the shares are indicative. The two departures leave different traces. Ideology works in one direction. It sustains patriarchy, never equality, and the ideological ruler cultivates the analogy even in egalitarian societies (21% of draws at $\mu = 1$, against 2% at $\mu = 0$), in an attempt to bring patriarchy back; the two configurations still do not coexist, since the ruler selects one. Tradition works in both directions. It sustains whichever configuration the analogy supports, and it brings back coexistence (10% of draws at $\tau = 0.75$), hence a role for history. In every case, power cultivates the analogy more where the society ends patriarchal (26 to 39% of draws) than where it ends egalitarian (2 to 21%).

The answer to the question we started from therefore has two parts. The patriarchal side has more interest in the analogy because the analogy transfers trust from those who behave better than the authority people see to those who behave worse, and in a patriarchal society with trusted fathers that trust is there to be borrowed by the ruler. But the analogy that power cultivates for this reason does not sustain patriarchy. It sustains it only when power wants patriarchy for itself, or when the analogy belongs to a tradition power did not make, and then only where fathers do not deserve trust on their own but the order around them does.

7 What is specific to the analogy, and how to test it

A persistent bias. Under the analogy the expectation about husbands is wrong by a fixed amount, which no amount of evidence corrects. This separates it from correctly specified learning.

Proposition 7 (A persistent bias)**.** *Under the coarse partition,* $e_F - \pi_F = [f_S(\pi_S - \pi_F) + f_O(\pi_O - \pi_F)]/\Sigma$*, which depends on relative exposures and not on how much evidence people hold. By contrast, let people believe that the conducts* (π_F, π_S, π_O) *are drawn from a joint normal prior with a positive definite covariance, and observe* Kf_k *independent encounters with holders of kind* k*. Their posterior expectation about husbands puts on the ruler's record a weight that tends to zero both as* $K \to 0$ *and as* $K \to \infty$*. If instead the prior forces one common type, the weight tends to* f_S/Σ*, the analogy weight, as* $K \to \infty$*.*

In the baseline at $(\beta, x) = (0.2, 0)$ the bias is -0.195: husbands are judged worse than they are, because rulers and other authorities are worse. Table 2 gives the Bayesian weight for a prior with correlation 0.6 across holders, and for one that forces a common type. In the first case the weight rises, since with little evidence the prior dominates, then falls to zero, since direct evidence on husbands eventually swamps the indirect evidence from rulers. In the second it converges to the analogy weight, 0.211. The analogy is thus the steady state of a learner who takes all holders of authority to share one type: a Berk–Nash equilibrium (Esponda and Pouzo, 2016), and the learning reading that Jehiel (2022) gives of his concept. The spillover that does not fade is the signature of this misspecification, not of analogy as against Bayesian inference. In the same reading, the ruler's technology is a technology of labels. The paternal metaphor files the state's acts under the heading people use for fathers, and Locke's reply to Filmer restores separate headings.

evidence K	1	10	100	1000
Bayesian, correlated types ($\rho = 0.6$)	0.030	0.083	0.032	0.004
Bayesian, one common type ($\rho = 1$)	0.049	0.158	0.204	0.210
Analogy	0.211	0.211	0.211	0.211

Table 2: Weight of the ruler's record in the expectation about husbands, baseline at $(\beta, x) = (0.2, 0)$. Bayesian: normal prior with variance 0.04 and correlation ρ across holders, noise variance 0.25, Kf_k observations of kind k.

Discourse. Section 5 predicts who uses which metaphor. The family metaphor should be more frequent in patriarchal societies, and in the mouths of rulers whose conduct is worse than that of fathers; rare among rulers with a good record of their own, who should insist on being judged on it; and absent where fathers are worse than the ruler, since he never pools with authorities below his class (Proposition 1). Where another authority is better regarded and more present than fathers, the metaphor should change: the ruler as shepherd where the clergy is trusted, as head of the firm where employers are. Where a society has become egalitarian, continued family rhetoric from power is a sign of ideology rather than of authority-seeking, since an authority-seeking ruler drops it there. Family metaphors in parliamentary and presidential speech can be counted by period, speaker and country, and matched to survey measures of trust in each kind of authority. One caution. Political speech mostly casts the state as a parent and citizens as children, while our family is a couple. The model extends to authority between generations, since what it needs is a holder of authority at home, but the measurement of husbands' conduct would then be replaced by that of parents'.

A survey experiment. Information that lowers the believed benevolence of one holder should lower the expectation about the other (Proposition 2). In the baseline, news that lowers the ruler's believed benevolence by 0.2 lowers the expected benevolence of a husband with authority from 0.60 to 0.56 in a patriarchal society; news about husbands lowers the expectation about the ruler from 0.60 to 0.50. Three arms: news about political abuse, news about domestic abuse, neutral news. Two outcomes: the acceptance of a husband's authority and the support for a strong leader. We predict that both cross effects are negative; that, while patriarchy is common, domestic news moves politics more than political news moves the family; and that neither effect fades with the experience respondents have of both spheres, holding their relative exposure fixed. The last prediction separates one type for all authority from correlated types: under the latter the cross effects shrink with age, years of marriage, or exposure to state authority. The feedback-spillover experiment of Huck et al. (2011), where analogy-based expectations emerge when feedback is coarse, is the template.

8 Final remarks

We have proposed a theory in which people judge the authority they meet by the authority they know. Husbands and rulers are pooled in one expectation, right on average and wrong case by case, and any such pooling transfers reputation from the better authorities to the worse. This

is why a ruler worse than fathers cultivates the family metaphor, and why he does so where patriarchy is common. But a ruler who chooses his metaphor joins only classes better than himself, so whoever he borrows from loses trust. The metaphor he cultivates can discredit fathers and never launder them: it rides on patriarchy without sustaining it. It sustains patriarchy only in the hands of a ruler who wants patriarchy, or of a tradition he did not make, and then only where fathers do not deserve trust on their own but the order around them does.

The theory holds fixed three things that are not. Conduct is a primitive, except for the ruler's; husbands' conduct plausibly responds to wives' outside options, which egalitarian households raise, and this would add a second force against self-concealment, as the selection of the husbands who remain already does in our checks. The expectation is single; wives are subject to husbands' authority while husbands exercise it, and separate exposures for women and men would give a gender gap in trust of strong rulers, largest where patriarchy is common. The limitation that matters most is the third: the analogy is symmetric. People who file the king under "father" also file fathers under "those in charge". If a ruler could make people judge him by fathers without making them judge fathers by him, the metaphor would cost families nothing, and the paper would have nothing to say about them. That the metaphor travels both ways is what the experiment above checks.

Whether a people trusts those in charge is learned at home before it is learned in the state. That is what the paternal metaphor exploits, and why it is worth more to a king than to a father.

A The joint system

This appendix completes Section 4 with four things: the share at which trust equals the threshold, the sharpness of choices that two family configurations require, the dynamics of the joint system, and the ruler who chooses his conduct.

Proposition 8 (Balance share)**.** *Let $\pi_F > \pi^*$. Then $\hat{\pi}(\beta, x) < \pi^*$ if and only if $x > \tilde{x}(\beta)$, where*

$$\tilde{x}(\beta) = 1 - \frac{\phi_S(1-\beta)\,(\pi^* - \pi_S) + \nu\,(\pi^* - \pi_O)}{\phi_F\,(\pi_F - \pi^*)}. \tag{10}$$

Moreover $\partial\tilde{x}/\partial\pi_S > 0$, and $\partial\tilde{x}/\partial\beta$ has the sign of $\pi^* - \pi_S$. If $\pi_F < \pi^*$, the comparison reverses.

The numerator collects the distrust produced outside the family, the denominator the trust that patriarchal husbands produce; new couples turn to equality when the first outweighs what remains of the second. In the baseline at $\beta = 0.2$, $\tilde{x} = -0.06$, 0.42 and 0.90 for $\pi_S = 0.2$, 0.5 and 0.8.

Proposition 9 (Families: coexistence)**.** *Fix β and write $h_\beta(x) = \Lambda(\kappa k_F(\pi^* - \hat{\pi}(\beta, x)))$ for the share of new couples who choose equality.*

(i) If $\pi_F > A(\beta)$, then h_β is increasing, and, generically, stable and unstable steady states of (4) alternate, the lowest and the highest being stable.

(ii) If in addition $\hat{\pi}(\beta,0) > \pi^* > \hat{\pi}(\beta,1)$*, which implies* $\pi_F > \pi^*$ *and* $\tilde{x}(\beta) \in (0,1)$*, then for every*

$$\kappa > \bar{\kappa}(\beta) = \frac{1}{k_F} \max \left\{ \frac{\ln(2/\tilde{x} - 1)}{\hat{\pi}(\beta, \tilde{x}/2) - \pi^*},\ \frac{\ln\big((1+\tilde{x})/(1-\tilde{x})\big)}{\pi^* - \hat{\pi}(\beta, (1+\tilde{x})/2)} \right\}$$

there are at least three steady states: a stable one in $(0, \tilde{x}/2)$*, an unstable one in* $(\tilde{x}/2, (1+\tilde{x})/2)$ *and a stable one in* $((1+\tilde{x})/2, 1)$*. As* $\kappa \to \infty$*, the stable ones converge to* 0 *and* 1 *and every unstable one to* $\tilde{x}(\beta)$*.*

(iii) If $\pi_F \leq A(\beta)$*, then* h_β *is nonincreasing and the family dynamics has a unique steady state, which attracts every path.*

The threshold is sufficient but tight: in the baseline at $\beta = \beta_m$ it gives $\bar{\kappa} = 19.3$, against an exact threshold of 19.1 found by computing the steady states.

The joint system. Let now both families and politics move. The two cross effects are

$$\frac{\partial \dot{x}}{\partial \beta} = \lambda \Lambda' \kappa k_F \frac{\phi_S(\pi_S - \hat{\pi})}{\Sigma}, \qquad \frac{\partial \dot{\beta}}{\partial x} = \eta\, s \frac{\phi_F(\pi_F - \hat{\pi})}{\Sigma} \quad \text{(zero where the target } b(\hat{\pi}) \text{ is clipped).}$$

Their signs record whether the ruler and the fathers are better or worse than the class average. When both are better, more participation hides a good ruler and more equality hides good fathers; both lower trust, and each change pushes the other along. When both are worse, the same logic runs in reverse. When fathers are better and the ruler worse, the case of the paternal king, the loop is negative: more equality hides good fathers, trust falls, participation rises, which hides the bad ruler, trust rises again, and equality loses ground. A negative loop can produce oscillations.

Theorem 3 (The joint system)**.** *Write* $\theta_F = \phi_F(\pi_F - \hat{\pi})/\Sigma$*,* $\theta_S = \phi_S(\hat{\pi} - \pi_S)/\Sigma$ *and* $\gamma = \kappa k_F \Lambda'$ *at a steady state.*

(i) (Reinforcement.) If $\hat{\pi} \leq \min(\pi_F, \pi_S)$ *at the four corners of the state space, then every path converges to a steady state. There is a largest steady state* z_E*, the egalitarian configuration when one exists, and for every* $\hat{\beta}$ *the shares* x *from which the path from* $(\hat{\beta}, x)$ *converges to it form an upper interval with lower end* $x_c(\hat{\beta})$*, the* tipping curve*, with* $x_c(\hat{\beta}) = 1$ *if the set is empty. The tipping curve is nonincreasing.*

(ii) If $\hat{\pi} \geq \max(\pi_F, \pi_S)$ *at the four corners, every path converges to a steady state.*

(iii) (The paternal king.) Let $\pi_S < \hat{\pi} < \pi_F$ *at a steady state where participation is interior, so that* $\theta_F, \theta_S > 0$*. The steady state is a saddle if* $\gamma\theta_F > 1 + s\theta_S$*. It attracts nearby paths if* $\gamma\theta_F < 1 + s\theta_S$ *and* $\lambda(\gamma\theta_F - 1) < \eta(1 + s\theta_S)$*, which holds in particular when* $\lambda \leq \eta$*. If* $1 < \gamma\theta_F < 1 + s\theta_S$ *and* $\lambda(\gamma\theta_F - 1) > \eta(1 + s\theta_S)$*, it repels nearby paths.*

The condition of part (i) requires both the ruler and the fathers to behave better than the class average wherever society is, which holds when the other authorities are the bad examples. Part (iii) is the case of the baseline. When households form and dissolve faster than politics

adjusts, the steady state between the two configurations can repel paths, and society can cycle: an example with $\lambda = 6.7$ and $\eta = 1$ cycles between $x = 0.24$ and $x = 0.85$ from three distant starts. Under condition (6) and with $\lambda \leq \eta$, as in all our sweeps, no steady state repels paths, whatever the ranking of conduct, and in 600 random draws integrated from 369 starts each, every path converged (Appendix C). With the exposure weights frozen at some state, every initial condition of the baseline ends at the same point, and if expectations adjust slowly toward the class average instead of instantly, outcomes do not change: the dynamic is a dynamic of exposure.

Proposition 10 (The ruler's conduct under the analogy)**.** *Let condition (6) hold, and let the ruler choose his benevolence $e = \pi_S$ to maximize $a(1-\beta) - \frac{c}{2}e^2$, where $\beta = \beta_m - s(\hat{\pi} - \pi^*)$ is the interior political steady state at the current x and $\hat{\pi}$ is evaluated at $\pi_S = e$. His first-order condition is*

$$e^* = \frac{a}{c} \cdot s \cdot \frac{f_S/\Sigma}{1 - s\,\phi_S\,(e^* - \hat{\pi})/\Sigma}.$$

Without the analogy the whole fraction is replaced by 1.

In the baseline at the patriarchal configuration the ruler chooses $e = 0.25$ under the analogy against 0.67 under the benchmark, the conduct at which participation reaches its lower bound, and his payoff is higher by 0.033; at the egalitarian configuration the analogy lowers his payoff by 0.128. This reads beliefs at the steady state: in a one-shot analogy-based equilibrium, expectations would not react to a deviation.

B Proofs

Throughout, $N = f_F\pi_F + f_S\pi_S + f_O\pi_O$, so that $\hat{\pi} = N/\Sigma$.

Proof of Lemma 1. $\sum_k f_k(e_k - \pi_k) = (fW - f)\cdot\pi$, which vanishes for every π if and only if $fW = f$ (the unit vectors are admissible conduct vectors). For a partition, $w_{kj} = f_j \mathbf{1}\{j \sim k\}/F_{C(k)}$ and $\sum_k f_k w_{kj} = f_j \sum_{k \in C(j)} f_k/F_{C(j)} = f_j$; consistency is preserved by convex combinations. Under the coarse partition $\sum_k f_k(\hat{\pi} - \pi_k) = \Sigma\hat{\pi} - N = 0$, and the sign of G_k is that of $\hat{\pi} - \pi_k$. □

Proof of Proposition 1. (i) $e_S(C \cup \{j\}) = [(f_S + F_C)\, e_S(C) + f_j\pi_j]/(f_S + F_C + f_j)$ is a weighted average of $e_S(C)$ and π_j with positive weights, so it exceeds $e_S(C)$ iff $\pi_j > e_S(C)$. Sufficiency of the top-set property: let $\lambda = e_S(C)$ with $\pi_j > \lambda$ for $j \in C$ and $\pi_j \leq \lambda$ for $j \notin C$, so that $\sum_{j \in C \cup \{S\}} f_j(\pi_j - \lambda) = 0$. For any C',

$$\sum_{j \in C' \cup \{S\}} f_j(\pi_j - \lambda) = \sum_{j \in C' \setminus C} f_j(\pi_j - \lambda) - \sum_{j \in C \setminus C'} f_j(\pi_j - \lambda) \leq 0,$$

hence $e_S(C') \leq \lambda$. Necessity: if C maximizes e_S and some $j \notin C$ had $\pi_j > e_S(C)$, adding it would raise the average; if some $j \in C$ had $\pi_j < e_S(C)$, then $e_S(C)$, an average of $e_S(C \setminus \{j\})$ and π_j, would lie strictly below $e_S(C \setminus \{j\})$. So a maximizer satisfies $\{\pi_j > \lambda\} \subseteq C \subseteq \{\pi_j \geq \lambda\}$, and every pooled kind has $\pi_j \geq \lambda$, with equality only for kinds whose conduct equals the average. The greedy procedure produces such a set: starting from $\emptyset$ with $e_S = \pi_S$, each addition raises the

average, so $e_S(C) \geq \pi_S$, and leaves it strictly below the conduct just added; later additions, of kinds with conduct no higher, keep the average below it; when the procedure stops, the remaining kinds have conduct at or below the average. With a cost $\kappa_j \geq 0$ per pooled kind, a pooled kind with $\pi_j < e_S(C)$ could be dropped, raising the average and saving its cost, so at an optimum every pooled kind has $\pi_j \geq e_S(C)$; the top-set structure itself can fail, since a costly kind above the average may be left out. (ii) $\partial G/\partial \pi_F = f_F/(f_S+f_F) > 0$, $\partial G/\partial f_F = f_S(\pi_F-\pi_S)/(f_S+f_F)^2$ and $\partial G/\partial f_S = -f_F(\pi_F-\pi_S)/(f_S+f_F)^2$. (iii) $G_S(\{j\}) = f_j(\pi_j-\pi_S)/(f_S+f_j)$ is increasing in f_j when $\pi_j > \pi_S$ and increasing in π_j. □

Proof of Proposition 2. $\partial\hat{\pi}/\partial\pi_S = f_S/\Sigma$ and $\partial\hat{\pi}/\partial\pi_F = f_F/\Sigma$; their ratio is f_F/f_S. For the exposures, $\partial\hat{\pi}/\partial x = (\partial_x N - \hat{\pi}\,\partial_x\Sigma)/\Sigma = -\phi_F(\pi_F-\hat{\pi})/\Sigma$, and likewise for β. Write $\hat{\pi} = [f_F\pi_F + (f_S+\nu)A(\beta)]/\Sigma$. Then $\pi_F - \hat{\pi} = (f_S+\nu)(\pi_F - A(\beta))/\Sigma$, whose sign does not depend on x; hence $\partial\hat{\pi}/\partial x$ has the constant sign of $A(\beta)-\pi_F$. The argument for β is the same with the average of fathers and other authorities, which does not depend on β. Under the benchmark $e_F = \pi_F$ and $e_S = \pi_S$ do not depend on the other sphere's conduct or on the state. □

Proof of Proposition 3. (i) $T_x(\beta) = b(\hat{\pi}(\beta,x))$ is continuous from $[b_\ell, b_h]$ into itself, so it has a fixed point. Where participation is interior, $T_x'(\beta) = b'(\hat{\pi})\,\partial_\beta\hat{\pi} = s\phi_S(\pi_S-\hat{\pi})/\Sigma$, and $T_x' = 0$ where b is at a bound. Since $|\pi_S - \hat{\pi}| \leq M - m$ and $\Sigma \geq \phi_F(1-x) + \phi_S(1-b_h) + \nu$, condition (6) gives $|T_x'| < 1$ everywhere, so $T_x(\beta) - \beta$ is strictly decreasing and has a unique zero, and $\dot{\beta} = \eta[T_x(\beta)-\beta]$ converges to it from every start. (ii) is Proposition 9. (iii) In the sharp-choice limit (4) becomes $\dot{x} = \lambda[\mathbf{1}\{\hat{\pi} < \pi^*\} - x]$. At $(\beta_P, 0)$ all new couples choose patriarchy if and only if $\hat{\pi}_P > \pi^*$, and then $x = 0$ is locally stable, participation converging to β_P by (i); if $\hat{\pi}_P < \pi^*$ every new couple chooses equality and x leaves 0. The egalitarian configuration is symmetric. If both are stable, $\hat{\pi}_E < \pi^* < \hat{\pi}_P$ and b is nonincreasing, so $\beta_E = b(\hat{\pi}_E) \geq b(\hat{\pi}_P) = \beta_P$. Now suppose $\pi_F \leq \pi^*$. Since $\hat{\pi}_P$ is a weighted average of π_F and $A(\beta_P)$ and exceeds $\pi^* \geq \pi_F$, $A(\beta_P) > \hat{\pi}_P > \pi^* > \hat{\pi}_E = A(\beta_E)$. As β rises, $A(\beta)$ moves toward π_O, monotonically; so $A(\beta_P) > A(\beta_E)$ with $\beta_E \geq \beta_P$ requires $\beta_E > \beta_P$ and $\pi_S > \pi_O$. Under the benchmark, $\dot{x} = \lambda[\mathbf{1}\{\pi_F < \pi^*\} - x]$ converges to 0 or to 1 from every start. □

Proof of Proposition 5. Without the analogy new couples choose patriarchy if and only if $\pi_F > \pi^*$, so a patriarchal configuration exists without the analogy if and only if $\pi_F > \pi^*$, and an egalitarian one if and only if $\pi_F < \pi^*$. With the analogy, by Proposition 3(iii), the patriarchal corner is stable if and only if $\hat{\pi}_P > \pi^*$ and the egalitarian corner if and only if $\hat{\pi}_E < \pi^*$. Combining gives the four cases. In case (iv), $\hat{\pi}_P$ is a weighted average of $\pi_F < \pi^*$ and $A(\beta_P)$ and exceeds π^*, so $A(\beta_P) > \pi_F$; in case (iii), $A(\beta_E) = \hat{\pi}_E > \pi^* > \pi_F$. In both, Proposition 9(iii) applies at the participation of the configuration. The corner forms of the conditions follow from $\Sigma(\hat{\pi} - \pi^*) = f_F(\pi_F - \pi^*) + f_S(\pi_S - \pi^*) + \nu(\pi_O - \pi^*)$ with $f_F = \phi_F$ at $x = 0$ and $f_F = 0$ at $x = 1$. □

Proof of Proposition 4. Let $T_x(\beta) = b(\hat{\pi}(\beta,x))$. Under (6), $T_x(\beta) - \beta$ is strictly decreasing and vanishes at the political steady state β_x of the corner (proof of Proposition 3(i)). Since $b_\ell < \beta_m < b_h$ and $b(\pi^*) = \beta_m$, $b(e) < \beta_m$ if and only if $e > \pi^*$. Hence $\hat{\pi}(\beta_x, x) > \pi^*$ iff $\beta_x = T_x(\beta_x) < \beta_m$ iff $T_x(\beta_m) < \beta_m$ iff $\hat{\pi}(\beta_m, x) > \pi^*$, and likewise with the inequalities reversed. Now $\Sigma(\beta_m, x)\,[\hat{\pi}(\beta_m, x) - \pi^*] = \phi_F(1-x)(\pi_F - \pi^*) - \Delta_O$, which is $\Delta_F - \Delta_O$ at $x = 0$

and $-\Delta_O$ at $x = 1$. With Proposition 3(iii) this gives the two stability conditions and the four regions, and with Proposition 5 the four cases.

Let $\Delta_F < \Delta_O < 0$. For every κ a steady state exists, by Brouwer's theorem applied to $(\beta, x) \mapsto (b(\hat{\pi}), \Lambda(\kappa k_F(\pi^* - \hat{\pi})))$. Let (β_κ, x_κ) be steady states with $\kappa \to \infty$ and (β, x) a limit point. If $x = 0$, β is the political steady state of the patriarchal corner, and $x_\kappa = \Lambda(\kappa k_F(\pi^* - \hat{\pi})) \to 0$ requires $\hat{\pi}(\beta, 0) \geq \pi^*$, which the first part excludes; likewise $x \neq 1$. So $x \in (0, 1)$, which requires $\hat{\pi} = \pi^*$ in the limit, hence $\beta = b(\pi^*) = \beta_m$ and $x = \tilde{x}(\beta_m) = 1 - \Delta_O/\Delta_F$ by Proposition 8. Participation is interior there, and the Jacobian computed in the proof of Theorem 3 gives $\det J = \lambda\eta(1 + s\theta_S - \gamma\theta_F)$ and $\operatorname{tr} J = \lambda(\gamma\theta_F - 1) - \eta(1 + s\theta_S)$, with $\theta_F \to \phi_F(\pi_F - \pi^*)/\Sigma < 0$ and $\gamma = \kappa k_F x_\kappa(1 - x_\kappa) \to \infty$. So $\det J \to +\infty$ and $\operatorname{tr} J \to -\infty$, and the steady state is locally stable for κ large. □

Proof of Theorem 1. (i) Let (C, ω) be optimal with $\omega > 0$, and write $e_S(\omega, C) = (1 - \omega)\pi_S + \omega e_S(C)$. If $e_S(C) \leq \pi_S$, then $e_S(\omega, C) \leq \pi_S = e_S(0, C)$, so authority at ω is no higher than at $\omega = 0$ while the cost is strictly higher, since $c_\omega > 0$: a contradiction. Hence $e_S(C) > \pi_S$. Suppose some $j \in C$ has $\pi_j < e_S(C)$, and let $C' = C \setminus \{j\}$. Since $e_S(C)$ is a weighted average of $e_S(C')$ and π_j, $e_S(C') > e_S(C) > \pi_S$. Put $\omega' = \omega\,(e_S(C) - \pi_S)/(e_S(C') - \pi_S) \in (0, \omega)$. Then $e_S(\omega', C') = e_S(\omega, C)$: the same expectation about the ruler, hence the same authority whether or not participation is at a bound, at a strictly lower cost of intensity and a no higher cost of metaphors. This contradicts optimality, so $\pi_j \geq e_S(C)$ for every $j \in C$. If $F \in C$, then $e_F = (1 - \omega)\pi_F + \omega e_S(C) \leq \pi_F$ for every $\omega \in [0, 1]$. (ii) If $F \notin C$ or $\omega = 0$, $e_F = \pi_F < \pi^*$; otherwise $e_F \leq \pi_F < \pi^*$ by (i). In both cases new couples choose equality. (iii) The first statement is the definition together with the observation that $e_F \neq \pi_F$ requires $F \in C$ and $\omega > 0$, and that $e_F < \pi^* < \pi_F$ then requires $e_S(C) < \pi^*$. For the egalitarian configuration, $x = 1$ gives $f_F = 0$, so $e_S(C)$ does not depend on whether $F \in C$. Self-confirmation requires $e_F < \pi^* < \pi_F$, hence $F \in C$ and $\omega > 0$; by (i), $e_S(C) > \pi_S$, so some $j \in C$ has $f_j > 0$ and $\pi_j > \pi_S$, and at $x = 1$ only O qualifies. Since including F changes neither $e_S(C)$ nor authority and costs κ_F, an optimum includes it only if $\kappa_F = 0$. □

Proof of Lemma 2. Write $\underline{e} = \pi^* - (b_h - \beta_m)/s$ and $d = \hat{\pi} - \pi_S$. Authority is $1 - b(e) = 1 - b_h + s\min\{(e - \underline{e})^+, \bar{e} - \underline{e}\}$, nondecreasing in e, and $e_S(\omega) = \pi_S + \omega d$. If $d \leq 0$, raising ω lowers e_S weakly and costs, so $\omega^* = 0$. If $d > 0$, authority has slope at most sd in ω while the cost has slope $c_\omega\omega$, so the objective decreases beyond sd/c_ω. Without a ceiling, $\pi_S \geq \underline{e}$, and the objective is $s\min\{\omega d, (\bar{e} - \pi_S)^+\} - \frac{c_\omega}{2}\omega^2$ up to a constant: strictly concave, maximized over $[0, 1]$ at the stated point. □

The ceiling. With a ceiling, $\pi_S < \underline{e}$, let $g = \underline{e} - \pi_S$, and let $w^\circ(d)$ and $W(d)$ be the maximizer and the maximum of $s\min\{\omega d, (\bar{e} - \pi_S)^+\} - \frac{c_\omega}{2}\omega^2$ over $[0, 1]$. For every ω the gain over $\omega = 0$ is $\max\{s\min\{\omega d, \bar{e} - \pi_S\} - sg, 0\} - \frac{c_\omega}{2}\omega^2$, so the ruler uses the analogy if and only if $W(d) > sg$, and then $\omega^* = w^\circ(d)$. Since W is continuous, nondecreasing, zero at $d = 0$ and tends to $s(\bar{e} - \pi_S) > sg$, he uses it if and only if d exceeds a threshold $d_J > 0$, and his weight jumps there. At a fixed corner the map $\beta \mapsto b(e_S)$ is still nonincreasing, but it jumps down where the class average equals $\pi_S + d_J$, and a configuration fails to exist when the diagonal passes through

the jump. Where configurations exist, parts (ii) to (iv) of Theorem 2 hold with "the ruler uses the analogy" replaced by "$W(\hat{\pi} - \pi_S) > sg$".

Proof of Theorem 2. At the corner x write $c_x = \phi_F(1-x)+\nu$ and $A_x = [\phi_F(1-x)\pi_F+\nu\pi_O]/c_x$, so that

$$\hat{\pi}(\beta, x) - \pi_S = \frac{c_x\,(A_x - \pi_S)}{c_x + \phi_S(1-\beta)}.$$

(i) If $\pi_S \geq A_x$, then $\hat{\pi} \leq \pi_S$ at the corner, $\omega^* = 0$, the configuration is $(b(\pi_S), 0)$, ω decays to zero and β converges to $b(\pi_S)$. If $\pi_S < A_x$, the class average $p_x(\beta)$ is increasing in β, and by Lemma 2 trust in the ruler at his best weight, $\hat{E}(p) = \pi_S + \min\{d,\, sd^2/c_\omega,\, (\bar{e} - \pi_S)^+\}$ with $d = p - \pi_S$, is continuous and nondecreasing in p. So $\mathcal{T}(\beta) = b(\hat{E}(p_x(\beta)))$ is a continuous, nonincreasing map of $[b_\ell, b_h]$ into itself: it has exactly one fixed point. At the corner, the field $(\dot{\beta}, \dot{\omega}) = (\eta[b(\pi_S + \omega(p_x(\beta) - \pi_S)) - \beta],\ \zeta[\omega^*(p_x(\beta)) - \omega])$ is Lipschitz on the forward-invariant rectangle $[b_\ell, b_h] \times [0, 1]$, and almost everywhere its divergence is $\eta[b'(e_S)\,\omega\,p_x'(\beta) - 1] - \zeta \leq -(\eta + \zeta)$. A closed orbit, or a closed curve made of orbits, would enclose a region on which the divergence integrates to a negative number while the flux through its boundary is zero (the divergence theorem, applied to a mollification of the field); so there is none. By the Poincaré–Bendixson theorem (Teschl, 2012, Theorem 7.16), whose proof uses only continuity of the field and uniqueness of solutions, the limit set of every path is then the unique steady state.

(ii) By Lemma 2, $\omega > 0$ at the configuration if and only if $\hat{\pi} > \pi_S$ and $\pi_S < \bar{e}$, and by the display $\hat{\pi} > \pi_S$ if and only if $A_x > \pi_S$. Where $\omega = 0$, $e_F = \pi_F$. If $\pi_S \geq \max\{A_P, \pi_O\}$, $\Sigma(\hat{\pi} - \pi_S) = \phi_F(1-x)(\pi_F - \pi_S) + \nu(\pi_O - \pi_S)$ is affine in x and nonpositive at both ends, so $\hat{\pi} \leq \pi_S$ at every state, $\omega^* \equiv 0$, $\omega \to 0$, $e_S \to \pi_S$ and $e_F \to \pi_F$; then every new couple eventually chooses the form that π_F supports.

(iii) $e_F < \pi^*$ if and only if $\omega(\pi_F - \hat{\pi}) > \pi_F - \pi^*$. This needs $\hat{\pi} < \pi^*$ and $\omega > 0$, hence $\pi_S < \hat{\pi}$. Then $\hat{\pi} < \pi^* < \bar{e}$, so the third term of ω^* exceeds one and $\omega = \min\{1, s(\hat{\pi} - \pi_S)/c_\omega\}$. With $r = (\pi_F - \pi^*)/(\pi_F - \hat{\pi}) \in (0, 1)$, $\omega > r$ if and only if $s(\hat{\pi} - \pi_S)/c_\omega > r$, which is the inequality defining I_D. Its left side, $s(p - \pi_S)(\pi_F - p)$, is a concave quadratic with maximum $s(\pi_F - \pi_S)^2/4$, so I_D is an interval, empty under the stated bound or when $\pi_S \geq \pi^*$. At the patriarchal corner $e_F < \pi^*$ is the failure of self-confirmation; at the egalitarian corner it is self-confirmation.

(iv) $e_F > \pi^*$ if and only if $\omega(\hat{\pi} - \pi_F) > \pi^* - \pi_F$. This needs $\hat{\pi} > \pi^*$ and $\omega > 0$, hence $\pi_S < \min\{\hat{\pi}, \bar{e}\}$. With $r = (\pi^* - \pi_F)/(\hat{\pi} - \pi_F) \in (0, 1)$, $\omega > r$ if and only if $s(\hat{\pi} - \pi_S)/c_\omega > r$ and $(\bar{e} - \pi_S)/(\hat{\pi} - \pi_S) > r$, which are the two inequalities defining I_L. The first holds above the larger root of a convex quadratic and the second is linear in p, so I_L is an interval or empty, and it is empty if $\pi_S \geq \bar{e}$. Since $\hat{\pi}$ is an average of $\pi_F < \hat{\pi}$, $\pi_S < \hat{\pi}$ and π_O, it lies below π_O; and $s(p - \pi_S)(p - \pi_F)$ is increasing in p above $\max\{\pi_S, \pi_F\}$, which gives the bound on c_ω. □

Proof of Proposition 6. For every share x the argument of Theorem 2(i), with c_x and A_x evaluated at x, gives one configuration (β, ω) of politics and of the ruler's weight; let $p(x)$ be its class average. Writing $p = \hat{\pi}(\mathcal{B}(p), x)$ with $\mathcal{B}(p) = b(\hat{E}(p))$ nonincreasing, $p'(x) = \partial_x\hat{\pi}/(1 - \partial_\beta\hat{\pi}\,\mathcal{B}')$ wherever $p > \pi_S$, with a denominator at least one; so p decreases in x where $p < \pi_F$ and increases where $p > \pi_F$ (Proposition 2). Take steady states with $\kappa \to \infty$ and a limit point. As in the proof of Proposition 4, its share is neither 0 nor 1, since neither corner is self-confirming; so $e_F \to \pi^*$ and the class average tends to an end of I_D (resp. I_L). If $\pi_F > \pi^*$: $p(0) \in I_D$, $p(0) < \pi^* < \pi_F$,

so p decreases from $p(0)$, never reaches the upper end of I_D, and crosses the lower end p_- once, since $p(1) \notin I_D$. If $\pi_F < \pi^*$: $p(1) \in I_L$ lies above $\max\{\pi_F, \pi_S\}$, so p increases up to $p(1)$, never reaches the upper end, and crosses the lower end once, since $p(0) \notin I_L$. In both cases the limit state is unique: with $\bar{p}$ the lower end, $\bar{\omega} = (\pi^* - \pi_F)/(\bar{p} - \pi_F)$, $\bar{\beta} = b(\pi_S + \bar{\omega}(\bar{p} - \pi_S))$, and $\bar{x} \in (0,1)$ solves $\hat{\pi}(\bar{\beta}, \bar{x}) = \bar{p}$.

The Jacobian of (β, x, ω), where it exists, has diagonal

$$\eta\big[b'(e_S)\,\omega\,\partial_\beta\hat{\pi} - 1\big], \qquad \lambda\Big[\gamma\,\omega\,\frac{\phi_F(\pi_F - \hat{\pi})}{\Sigma} - 1\Big], \qquad -\zeta, \qquad \gamma = \kappa k_F \Lambda',$$

and $\gamma = \kappa k_F x(1-x)$ at a steady state. (i) Since $\omega(\pi_F - \hat{\pi}) = \pi_F - e_F \to \pi_F - \pi^* > 0$ and $\bar{x}(1-\bar{x}) > 0$, the middle term tends to $+\infty$ uniformly on a neighbourhood of the steady state that shrinks with κ, while the others are bounded; so for κ large the divergence of the field exceeds some $\delta > 0$ almost everywhere on that neighbourhood. If the steady state were stable, some small ball around it would stay inside that neighbourhood for all times, while by Liouville's formula its volume would grow at least like $e^{\delta t}$: a contradiction. Condition (9) makes the trace positive, using $b' \geq -s$ and $\omega(\hat{\pi} - \pi_S) = e_S - \pi_S$. (ii) At $\bar{p}$ the inequality $(\bar{e} - \pi_S)(p - \pi_F) > (\pi^* - \pi_F)(p - \pi_S)$ is strict, so the ruler's weight is $s(\bar{p} - \pi_S)/c_\omega$ or 1, and participation is interior. Write $J = J_0 + \gamma\lambda\,\mathbf{e}_x v^\top$ with $v = -\nabla e_F$. By the matrix determinant lemma, as $\gamma \to \infty$ one eigenvalue behaves like $\gamma\lambda v_x = -\gamma\lambda\,\omega\,\phi_F(\hat{\pi} - \pi_F)/\Sigma \to -\infty$, and the other two tend to those of the system restricted to $e_F = \pi^*$. On that surface $\hat{\pi} = P(\omega) = \pi_F + (\pi^* - \pi_F)/\omega$, and the restricted system $\dot{\omega} = \zeta[\omega^*(P(\omega)) - \omega]$, $\dot{\beta} = \eta[b(\pi_S + \omega(P(\omega) - \pi_S)) - \beta]$ is triangular, with eigenvalues $-\eta$ and $\zeta[\omega^{*\prime}(\bar{p})P'(\bar{\omega}) - 1] < 0$, since $P' < 0$ and ω^* is nondecreasing at $\bar{p}$. □

Proof of Proposition 7. $\hat{\pi} - \pi_F = [f_S(\pi_S - \pi_F) + f_O(\pi_O - \pi_F)]/\Sigma$ since the F term vanishes. For the Bayesian, let $\bar{y}$ be the vector of sample means of the three records, with $\bar{y} = \pi + \varepsilon$ and $\varepsilon \sim N(0, R_K)$, $R_K = \mathrm{diag}(\sigma^2/(Kf_k))$. With prior $\pi \sim N(\pi_0, V)$, the posterior mean is $\pi_0 + W_K(\bar{y} - \pi_0)$ with $W_K = V(V + R_K)^{-1}$. As $K \to \infty$, $R_K \to 0$ and $W_K \to VV^{-1} = I$, whose off-diagonal entries are zero. As $K \to 0$, $W_K = KV\big(KV + \mathrm{diag}(\sigma^2/f_k)\big)^{-1} \to 0$. Hence $w_{FS}(K) = (W_K)_{FS} \to 0$ at both ends. With one common type, $\pi_k = \theta$ for all k and $\theta \sim N(\theta_0, v)$; the posterior mean of θ puts on the ruler's sample mean the weight $(Kf_S/\sigma^2)/(1/v + K\Sigma/\sigma^2)$, which tends to f_S/Σ as $K \to \infty$. □

Proof of Proposition 8. $\pi^* - \hat{\pi}$ has the sign of $\Sigma(\pi^* - \hat{\pi}) = \phi_F(1-x)(\pi^* - \pi_F) + f_S(\pi^* - \pi_S) + \nu(\pi^* - \pi_O)$, which is affine in x with slope $\phi_F(\pi_F - \pi^*)$. When $\pi_F > \pi^*$ it is increasing in x and vanishes at $\tilde{x}$ given by (10), so $\hat{\pi} < \pi^*$ iff $x > \tilde{x}$. The derivatives are read off (10): $\partial\tilde{x}/\partial\pi_S = \phi_S(1-\beta)/[\phi_F(\pi_F - \pi^*)] > 0$ and $\partial\tilde{x}/\partial\beta = \phi_S(\pi^* - \pi_S)/[\phi_F(\pi_F - \pi^*)]$. When $\pi_F < \pi^*$ the affine function is decreasing in x, which reverses the comparison. □

Proof of Proposition 9. (i) If $\pi_F > A(\beta)$ then $\hat{\pi}$ is decreasing in x by Proposition 2, and $h_\beta = \Lambda(\kappa k_F(\pi^* - \hat{\pi}))$ is increasing. Steady states are the zeros of $g(x) = h_\beta(x) - x$, with $g(0) = h_\beta(0) > 0$ and $g(1) = h_\beta(1) - 1 < 0$. Since $\dot{x} = \lambda g(x)$, a zero at which g changes sign from $+$ to $-$ is stable and one at which it changes from $-$ to $+$ is unstable; sign changes alternate, and the first and last are from $+$ to $-$. Generically every zero is a sign change.

(ii) Under the additional condition, $\hat{\pi}(\beta,\cdot)$ is decreasing (which also gives $\pi_F > A(\beta)$) and crosses π^* at $\tilde{x} \in (0,1)$, so $D_1 = \hat{\pi}(\beta,\tilde{x}/2) - \pi^* > 0$ and $D_2 = \pi^* - \hat{\pi}(\beta,(1+\tilde{x})/2) > 0$. For $\kappa > \bar{\kappa}$, $\kappa k_F D_1 > \ln(2/\tilde{x} - 1)$, which is equivalent to $\Lambda(-\kappa k_F D_1) < \tilde{x}/2$, that is $g(\tilde{x}/2) < 0$; and $\kappa k_F D_2 > \ln((1+\tilde{x})/(1-\tilde{x}))$ is equivalent to $\Lambda(\kappa k_F D_2) > (1+\tilde{x})/2$, that is $g((1+\tilde{x})/2) > 0$. With $g(0) > 0$ and $g(1) < 0$, the intermediate value theorem gives zeros in $(0,\tilde{x}/2)$, in $(\tilde{x}/2,(1+\tilde{x})/2)$ and in $((1+\tilde{x})/2,1)$, with sign changes $+ \to -$, $- \to +$ and $+ \to -$ respectively, hence stable, unstable, stable. As $\kappa \to \infty$, $h_\beta(x) \to 0$ for $x < \tilde{x}$ and $\to 1$ for $x > \tilde{x}$ uniformly on compacts away from $\tilde{x}$, so any zero of g in $[0,\tilde{x}-\delta]$ tends to 0, any zero in $[\tilde{x}+\delta,1]$ tends to 1, and the remaining zeros lie within δ of $\tilde{x}$ for κ large.

(iii) If $\pi_F \le A(\beta)$ then $\hat{\pi}$ is nondecreasing in x and h_β nonincreasing, so g is strictly decreasing, has one zero, and $\dot{x} = \lambda g(x)$ converges to it. □

Proof of Theorem 3. The vector field $F(\beta,x) = (\eta[b(\hat{\pi}) - \beta],\ \lambda[h_\beta(x) - x])$ is Lipschitz on the compact, convex and forward-invariant state space $[b_\ell, b_h] \times [0,1]$. It is differentiable except on the kink of b, where $\partial\dot{\beta}/\partial x$ jumps between $\eta s \phi_F(\pi_F - \hat{\pi})/\Sigma$ and 0, without changing sign. The results below use only the sign of each component's dependence on the other variable (the Kamke condition), which this preserves.

By Proposition 2, $\hat{\pi}$ is monotone along every horizontal and every vertical segment of the state space, so its maximum and minimum are reached at corners. Hence the conditions of (i) and (ii) hold everywhere.

(i) If $\hat{\pi} \le \min(\pi_F,\pi_S)$ everywhere, both cross effects are nonnegative: the system is cooperative for the product order $(\beta,x) \le (\beta',x')$ iff $\beta \le \beta'$ and $x \le x'$. The flow is then monotone, $z \le z'$ implying $\varphi_t(z) \le \varphi_t(z')$ (Smith, 1995, Chapter 3). In a planar cooperative or competitive system every bounded forward orbit is eventually monotone in each coordinate, hence converges to a steady state (Hirsch, 1982). These results are usually stated for C^1 fields, but the argument uses only uniqueness of solutions and order preservation, which hold for Lipschitz fields satisfying the Kamke condition (Smith, 1995, Chapter 3). The path from the top corner $(b_h,1)$ is nonincreasing, since F points into the state space there, so it converges to a steady state z_E. Every steady state z satisfies $z \le (b_h,1)$, hence $z = \varphi_t(z) \le \varphi_t(b_h,1) \to z_E$: z_E is the largest steady state. If the path from z converges to z_E and $z \le z'$, the path from z' converges to a steady state $z'' \ge z_E$, hence $z'' = z_E$. The basin of z_E is therefore an upper set. At fixed $\hat{\beta}$ it is an upper interval in x, possibly empty. If $\beta < \beta'$, then $(\beta,x) \le (\beta',x)$, so the basin at β' contains the basin at β, and $x_c(\beta') \le x_c(\beta)$.

(ii) If $\hat{\pi} \ge \max(\pi_F,\pi_S)$ everywhere, both cross effects are nonpositive and the system is competitive; convergence follows as in (i).

(iii) With participation interior, $\partial\dot{\beta}/\partial\beta = \eta(s\phi_S(\pi_S - \hat{\pi})/\Sigma - 1) = -\eta(1 + s\theta_S)$, $\partial\dot{\beta}/\partial x = \eta s\theta_F$, $\partial\dot{x}/\partial\beta = -\lambda\gamma\theta_S$ and $\partial\dot{x}/\partial x = \lambda(\gamma\theta_F - 1)$. Hence

$$\det J = \lambda\eta(1 + s\theta_S - \gamma\theta_F), \qquad \operatorname{tr} J = \lambda(\gamma\theta_F - 1) - \eta(1 + s\theta_S).$$

If $\gamma\theta_F > 1 + s\theta_S$, $\det J < 0$ and the steady state is a saddle. If $\gamma\theta_F < 1 + s\theta_S$, $\det J > 0$, and the steady state attracts nearby paths iff $\operatorname{tr} J < 0$, that is iff $\lambda(\gamma\theta_F - 1) < \eta(1 + s\theta_S)$. When $\gamma\theta_F \le 1$ this always holds. When $1 < \gamma\theta_F < 1 + s\theta_S$, it holds if $\lambda \le \eta$, since then $\lambda(\gamma\theta_F - 1) < \lambda s\theta_S \le \eta s\theta_S < \eta(1 + s\theta_S)$; it fails if $\lambda(\gamma\theta_F - 1) > \eta(1 + s\theta_S)$, and the steady state

then repels nearby paths.

For the remark after the theorem: with $q = 1 + s\theta_S = 1 - T'_x$, condition (6) gives $q > 0$. A steady state repels nearby paths only if $\det J > 0$ and $\operatorname{tr} J > 0$, that is $\gamma\theta_F < q$ and $\lambda(\gamma\theta_F - 1) > \eta q$. If $q \leq 1$ the second fails since $\gamma\theta_F - 1 < 0$. If $q > 1$ it requires $\lambda(q-1) > \eta q$, hence $\lambda > \eta$. □

Proof of Proposition 10. Write $\beta(e) = \beta_m - s(\hat{\pi}(e) - \pi^*)$ with $\hat{\pi}$ depending on e directly, through $\pi_S = e$, and through β: $\mathrm{d}\hat{\pi}/\mathrm{d}e = f_S/\Sigma + (\partial\hat{\pi}/\partial\beta)(\mathrm{d}\beta/\mathrm{d}e) = f_S/\Sigma - \phi_S(e - \hat{\pi})/\Sigma \cdot (-s\,\mathrm{d}\hat{\pi}/\mathrm{d}e)$, so $\mathrm{d}\hat{\pi}/\mathrm{d}e = (f_S/\Sigma)/(1 - s\phi_S(e - \hat{\pi})/\Sigma)$. The first-order condition $a\,s\,\mathrm{d}\hat{\pi}/\mathrm{d}e = c\,e$ gives the formula. Without the analogy, $e_S = e$ and $\mathrm{d}\beta/\mathrm{d}e = -s$, so the factor is 1. □

C Numerical methods, sweeps and checks

Integration. Paths are computed by a linearly implicit Euler scheme with step 0.1 (step 0.05 when the ruler sets the analogy), which is stable when choices are sharp; it reproduces an adaptive stiff solver exactly for regimes and basins on 60 test draws. Baseline numbers use the adaptive solver.

Ranges. The sweeps draw $\pi_F \in [0.55, 0.95]$ (the regime sweep below) or $[0.2, 0.95]$ (Sections 3, 4 and 5), $\pi_S \in [0.1, 0.9]$, $\pi_O \in [0.1, 0.6]$, $\phi_F \in [0.5, 2]$, $\phi_S \in [0.2, 1]$, $\nu \in [0.05, 1]$, $\pi^* \in [0.4, 0.65]$, $k_F \in [1, 4]$, $s \in [0.5, 3]$, $\kappa \in [5, 100]$, $\lambda \in [0.1, 1]$, by Latin hypercube; the bounds on participation and β_m are those of the baseline. The ranges are wide on purpose. The shares reported describe what the model can produce, not the frequency of configurations in real societies.

What the model can produce. Over 20,000 parameter vectors with the analogy at full weight, two stable configurations occur in 28.6% of draws, a patriarchal configuration only in 33.2%, an egalitarian one only in 33.4%, a stable mixture in 4.1%, and two stable configurations one of which is a mixture in 0.7% (Figure 2). Below $\kappa k_F \approx 20$ there is a single outcome, usually a mixture. A Sobol decomposition of the variance of the coexistence indicator, on a design of base 2,048 (26,624 draws), gives total indices π^* 0.59, π_S 0.58, π_O 0.47, π_F 0.45, ν 0.34, ϕ_S 0.25, κ 0.25, ϕ_F 0.23, k_F 0.13, s 0.02, λ 0.01, first-order indices summing to 0.29. For the size of the egalitarian basin, the ruler's conduct has the second largest total index (0.43), ahead of husbands' own conduct (0.22). Bootstrap intervals over base samples are within ±0.04 of the totals. The political road to two configurations noted after Proposition 3 appears in none of these draws and in none of 400,000 draws aimed at it with the iteration that finds one political steady state per corner; an example is $\pi_F = 0.45$, $\pi_S = 0.9$, $\pi_O = 0.1$, $\phi_F = 0.1$, $\phi_S = 3$, $\nu = 1.8$, $\pi^* = 0.5$, $s = 6$ and $\kappa = 100$, with paths ending at $(\beta, x) = (0.15, 0)$ or $(0.5, 1)$ depending on where participation starts. At $\kappa = 50$, condition (7) agrees with the simulated outcome in 162 of 195 random draws; the others have κ below the threshold of Proposition 9(ii).

The four cases. Table 3 gives the frequency of each case of Proposition 5 in 50,000 draws with $\pi_F \in [0.2, 0.95]$, the agreement of the closed forms with the simulated outcome at $\kappa = 50$, and who gains in each. Fathers gain from the analogy when the ruler is good (cases iii and iv),

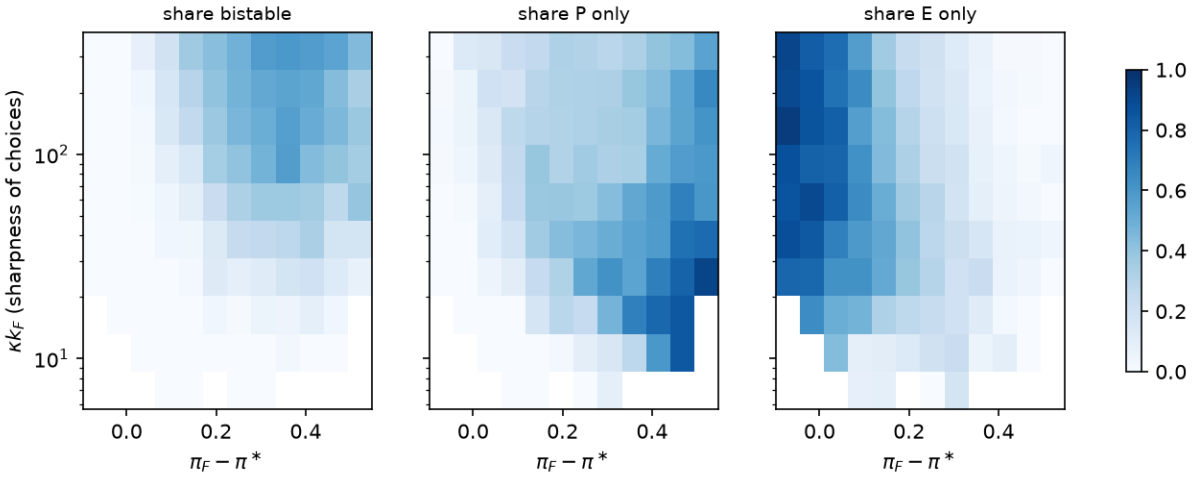


Figure 2: Share of each regime by cell of $(\pi_F - \pi^*,\ \kappa k_F)$ over the 20,000 draws with the analogy at full weight (cells with at least 10 draws).

case	share of draws	agreement with Prop. 5	fathers gain	ruler gains
(i) equality needs the analogy	36.8%	91%	0%	39%
(ii) patriarchy destroyed by it	17.3%	91%	0%	43%
(iii) equality destroyed by it	9.2%	97%	100%	8%
(iv) patriarchy needs the analogy	2.2%	99%	100%	2%

Table 3: The four cases over 50,000 draws with $\pi_F \in [0.2, 0.95]$. "Agreement" is the share of draws in which the closed form matches the simulated outcome at $\kappa = 50$. "Fathers gain" and "ruler gains" are the shares of draws in the case where $G_F > 0$ and $G_S > 0$ at the configuration concerned.

and the ruler gains when fathers are good (cases i and ii); over all draws, the ruler gains in 57% of patriarchal configurations and 38% of egalitarian ones. Patriarchy itself is not rare, a stable patriarchal configuration existing in 45% of draws; what is rare is a patriarchy that exists only thanks to the analogy. Part of the reason is in the ranges: other authorities deserve trust in 16% of draws only. In a symmetric world, where every conduct is drawn over $[0.2, 0.95]$ and the state is met as often as the family, case (iv) occurs in 13% of draws and case (i) in 22%; in a world where the state and other authorities are benevolent and fathers mediocre, case (iv) occurs in 36% and case (i) in 0.1%. Figure 3 maps the cases.

Checks of the formal results. The script `check_math.py` checks the results of Sections 4 and Appendix A on random draws. The transfer identity holds to 6×10^{-16} over 2,000 states. The monotonicity of $\hat{\pi}$ in x, with the sign of $A(\beta) - \pi_F$, holds in 2,000 of 2,000 draws. The uniqueness condition (6) holds in 1,934 of 2,000 draws, all with a unique political steady state. At $\kappa = 1.001\,\bar{\kappa}$, the three steady states of Proposition 9(ii) lie in the stated intervals in 784 of 793 draws; in the other nine the logistic choice rounds to one at $x = 1$ in floating point, so the root at the upper end is not detected. For Theorem 3, 300 draws with $\pi_S \leq \pi_O \leq \pi_F$ and 300

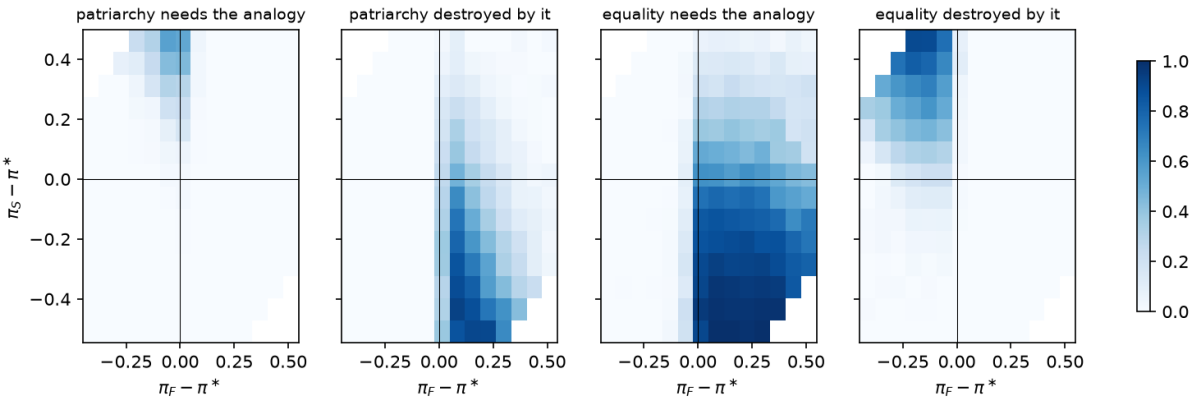


Figure 3: Share of draws in each case of Proposition 5, in the plane $(\pi_F - \pi^*,\ \pi_S - \pi^*)$ (cells with at least 15 draws).

unrestricted draws, all with $\lambda \leq \eta = 1$, were integrated from 369 starts each, with step 0.02 and horizon 650: every path converged, the egalitarian basin was an upper interval in x at every β, and in the 57 ordered draws reaching both configurations the tipping curve was nondecreasing, which is numerical, not a theorem. The limit cycle of part (iii) is reproduced with step 0.005. For parts (i) and (ii), 141 draws satisfying the condition of (i) and 19 satisfying that of (ii), with λ and η drawn in $[0.1, 10]$, were integrated with step 0.002: every path converged, every basin was an upper interval, and in the 20 draws of part (i) reaching both configurations the tipping curve was nonincreasing. For Theorem 2, in 200 random draws at $\kappa = 200$ the prediction of part (i) matches the simulated stability, from a start next to the configuration, of the patriarchal configuration in 197 and of the egalitarian one in 200.

The script `check_d2.py` checks Propositions 4 and 6, Lemma 2 and Theorem 2, on uniform draws in the ranges above and on draws where the bounds on participation and β_m are also random. The stability conditions of Proposition 4 hold without exception in 100,000 draws at each corner, and simulated paths at $\kappa k_F = 800$ agree with them in 2,000 of 2,000 draws at each corner; in the region where neither configuration is stable, the path ends within 0.017 of $(\beta_m, 1 - \Delta_O/\Delta_F)$. The formula of Lemma 2, and its extension to the ceiling, equal the maximizer found on a grid in 40,000 draws. Without a ceiling, a configuration exists at each corner in every draw, and paths of (β, ω) at a fixed corner reach it to 10^{-11} from four starts in 2,000 draws. With a ceiling (37% of draws), no configuration exists at the patriarchal corner in 0.6% of draws and at the egalitarian one in 0.07%, exactly where the criterion of the jump predicts. Over 250,109 draws without a ceiling, the events of parts (iii) and (iv) of Theorem 2 coincide with "the class average lies in the interval" in every draw (1,052 events); with a ceiling, after adding the condition that the analogy pays, in every one of 149,775 draws. Families follow fathers at both corners in 99.6% of the draws without a ceiling. With the class chosen among $\emptyset$, $\{F\}$, $\{O\}$, $\{F, O\}$, the patriarchal configuration fails with trusted fathers exactly when the average of the chosen class lies in I_D (2,393 cases in 400,000 draws). For Proposition 6, with cheap propaganda ($c_\omega \in [0.02, 0.4]$) and $\kappa k_F = 800$, among 400 draws with trusted fathers and neither configuration self-confirming, the steady state is unstable in 364 and the share keeps

moving on both paths from the corners in 362; at the draws' own sharpness, the steady state is unstable in 87 and the share keeps moving in all 87. Among 400 draws with mediocre fathers, the steady state is stable and the paths settle in all 400, within 0.05 of the limit state. From starts next to a self-confirming configuration, paths return to it in every case: 2,486 starts without a ceiling and 2,466 with one. Under the chosen class (script `check_d2_chosen.py`), with trusted fathers, no ceiling and $c_\omega \in [0.02, 0.4]$, neither configuration is self-confirming in 0.76% of 545,087 draws, against 0.95% under the given class; among the 232 such draws where both corners miss the threshold by more than 0.03, the share keeps moving from the patriarchal corner in 89 at $\kappa k_F = 800$ and in 15 at the draws' own sharpness.

The script `check_technology.py` checks the results on the analogy technology (Sections 3 and 5). Consistency and the transfer identity hold to 7×10^{-16} for random partitions and mixtures of partitions over 2,000 draws with two to six kinds, and fail for random symmetric similarity kernels in 2,000 of 2,000 draws, with an average bias of up to 0.30 in trust. The brute-force maximizer of $e_S(C)$ over all subsets equals the greedy top set in 3,000 of 3,000 draws, and no pooled kind has conduct below the class average, with or without per-kind costs. The monotonicities of Proposition 1(ii) hold in 20,000 draws. For Theorem 1, 20,000 draws in the ranges above, at the patriarchal corner, with the fixed point in (β, C, ω) computed by damped iteration, the class chosen among $\emptyset$, $\{F\}$, $\{O\}$, $\{F, O\}$ and ω on a grid of 201 levels with the quadratic cost, participation possibly at a bound, and ties broken toward the smaller class. The damped iteration converges in 98.4% of draws under the chosen class and 99.2% under the given one; the rest, in which it cycles, are excluded, leaving 19,558 draws. The numbers reported in Sections 3 and 5 are the share of draws with mediocre fathers in which the patriarchal corner is self-confirming under the chosen and under the given class, the share with trusted fathers in which it fails under each, and the metaphor shares; pooled fathers lose trust in every draw. At the egalitarian corner with ties broken toward including fathers, the corner is self-confirming with trusted fathers in 74 of 11,238 converged draws, all with $\pi_O > \pi_S$; with ties toward the smaller class, in none. The dynamics under the chosen class (`manip_class.py`) were integrated on 2,000 draws over the ranges of Section 5, from both starts, with the class chosen among $\emptyset$, $\{F\}$, $\{O\}$, $\{F, O\}$ and the weight by a global search at each instant, ties toward the smaller class, the class switching at once and the weight adjusting at speed ζ; the same draws were integrated under the given class. Of the 881 draws with mediocre fathers, none ends patriarchal under either technology; of the 1,119 with trusted fathers, 1.3% (patriarchal start) and 2.1% (egalitarian start) end with $x > 0.5$ under the chosen class, against 1.0 and 1.3% under the given one, most of them at a mixture with the metaphor kept; the end label (patriarchal, egalitarian or cycle) is the same under the two technologies in 99.05% of (draw, start) pairs; no draw cycles under the chosen class, against 0.5% under the given one; where society ends patriarchal the ruler's class is the family alone in 43% of cases, both in 1% and none in 56%, and where it ends egalitarian none in 92%, the clergy in 4%, the family in 2% and both in 2%. For the population reading of ω, the same corner computation with each kind of citizen choosing on her own expectation, so that the share of new couples choosing equality is $(1-\omega)\mathbf{1}\{\pi_F < \pi^*\} + \omega\mathbf{1}\{\hat{\pi} < \pi^*\}$, gives the comparison reported in Section 2: the end state follows the sign of $\pi_F - \pi^*$ in 99.7% of draws from either start under the average expectation and in 99.5 to 99.7% under the population reading; the population reading ends in a mixed society in 2.2 to 2.8% of draws against 0.7%;

the ruler keeps the analogy in 30% of patriarchal ends and 4% of egalitarian ends under both. Table 2 comes from the same script.

Robustness. The regime sweep was recomputed with a convex and a concave political rule, with probit instead of logit choices, with a political trust threshold different from the family's, and with selection of the husbands who remain (abused wives leave first, so π_F rises with x). The shares change by less than two points, except under selection, which makes patriarchy more resilient: the egalitarian basin shrinks from 0.36 to 0.30 among draws with two stable configurations. The results of Section 5 were recomputed with probit choices (identical) and with selection (coexistence still disappears; where fathers start below the threshold the outcome is split, since the remaining fathers improve as egalitarian households spread).

Replication. All numbers come from the scripts in `explore_jehiel/` of the replication folder: `jehiel_v2.py` (model), `check_jehiel3.py` and `check_jehiel4.py` (baseline results), `check_math.py`, `check_technology.py`, `check_d2.py` and `check_d2_chosen.py` (checks of the formal results), `sweep_fast.py`, `sweep_parallel.py` and `sweep_variants.py` (sweeps), `analyze_sweeps.py`, `night_base.py` and `analyze_base.py` (both rankings), `rarity.py`, `who_gains.py`, `manipulation.py`, `overnight_manip.py`, `analyze_manip.py`, `manip_variants.py`, `analyze_variants.py` and `manip_class.py` (simulations of Sections 5 and 6).

References

Alesina, A. and Giuliano, P. (2011). Family ties and political participation. *Journal of the European Economic Association*, 9(5):817–839.

Bisin, A. and Verdier, T. (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*, 97(2):298–319.

Bisin, A. and Verdier, T. (2024). On the joint evolution of culture and political institutions: Elites and civil society. *Journal of Political Economy*, 132(5):1485–1564.

Esponda, I. and Pouzo, D. (2016). Berk–Nash equilibrium: A framework for modeling agents with misspecified models. *Econometrica*, 84(3):1093–1130.

Ettinger, D. and Jehiel, P. (2010). A theory of deception. *American Economic Journal: Microeconomics*, 2(1):1–20.

Filmer, R. (1680). *Patriarcha, or the Natural Power of Kings*. London.

Guriev, S. and Treisman, D. (2020). A theory of informational autocracy. *Journal of Public Economics*, 186:104158.

Hirsch, M. W. (1982). Systems of differential equations which are competitive or cooperative I: Limit sets. *SIAM Journal on Mathematical Analysis*, 13(2):167–179.

Huck, S., Jehiel, P., and Rutter, T. (2011). Feedback spillover and analogy-based expectations: A multi-game experiment. *Games and Economic Behavior*, 71(2):351–365.

Jehiel, P. (2005). Analogy-based expectation equilibrium. *Journal of Economic Theory*, 123(2):81–104.

Jehiel, P. (2022). Analogy-based expectation equilibrium and related concepts: Theory, applications, and beyond. Working paper, Paris School of Economics, HAL halshs-03735680.

Jehiel, P. and Samuelson, L. (2012). Reputation with analogical reasoning. *Quarterly Journal of Economics*, 127(4):1927–1969.

Lakoff, G. (1996). *Moral Politics: What Conservatives Know That Liberals Don't*. University of Chicago Press, Chicago.

Locke, J. (1689). *Two Treatises of Government*. London.

Mohlin, E. (2014). Optimal categorization. *Journal of Economic Theory*, 152:356–381.

Mullainathan, S., Schwartzstein, J., and Shleifer, A. (2008). Coarse thinking and persuasion. *Quarterly Journal of Economics*, 123(2):577–619.

Schwartzstein, J. and Sunderam, A. (2021). Using models to persuade. *American Economic Review*, 111(1):276–323.

Smith, H. L. (1995). *Monotone Dynamical Systems: An Introduction to the Theory of Competitive and Cooperative Systems*. Mathematical Surveys and Monographs 41. American Mathematical Society, Providence, RI.

Steiner, J. and Stewart, C. (2008). Contagion through learning. *Theoretical Economics*, 3(4):431–458.

Teschl, G. (2012). *Ordinary Differential Equations and Dynamical Systems*. Graduate Studies in Mathematics 140. American Mathematical Society, Providence, RI.

Todd, E. (1985). *The Explanation of Ideology: Family Structures and Social Systems*. Basil Blackwell, Oxford. Translation of *La Troisième Planète*, Seuil, Paris, 1983.

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4):327–352.