Hierarchy and the Scaling of Trust

Susan Athey Emilio Calvano Saumitra Jha*

April 30, 2026

Abstract

From early human groups, to classical republics, to modern open source software communities, many societies scale cooperative economic activity through *hierarchies of trust*: a public, coarse, and slowly evolving rank order that regulates who is trusted, by how much, and by whom. We model such a *hierarchy* as a family of stationary sequential equilibria in which principals condition trust on the agent's seniority and agents condition cooperation and mobility on their continuation value. We show that as the number of tasks individuals can be entrusted to do for others increases, the maximum community size compatible with cooperation *scales* linearly. Further, we show that hierarchies enjoy *dynamic evolutionary stability*: active members strictly prefer the hierarchy to more egalitarian alternatives, and hierarchies may even lead to the breakdown of trust in these communities. We map the hierarchical equilibrium to key features of the early Roman Republic to explain its ability to scale and ability to absorb new members relative to its more internally egalitarian but exclusive rivals like Carthage and Athens, before describing how our model applies elsewhere, from the emergence of hierarchies in early human societies to open source software groups.

JEL codes: Z13, C73, D02, N40

* *athey@stanford.edu*; *emilio.calvano@luiss.it*; *saumitra@stanford.edu*. A previous version of this paper circulated as *A Theory of Community Formation and Social Hierarchy.* We would like to thank David Baron, Avinash Dixit, Avner Greif, Jim Fearon, David Laitin, Pedro Miranda, Ian Morris, Josiah Ober, Paul Seabright, Adam Szeidl and participants in the Berkeley development seminar, Harvard-MIT organizational economics seminar, Nobel Symposium on Foundations of Organization (Stockholm), 4eme JEC workshop (Paris), and across Stanford for helpful comments. Gleb Romanyuk provided much valued assistance. Athey thanks the Toulouse Network for Information Technology for support.

1 Introduction

From early human groups, to classical republics, to modern open source software communities, many societies have developed means to sustain and scale cooperative economic activity through *hierarchies of trust*: a public, coarse, and slowly evolving rank order that regulates who is trusted, by how much, and by whom. Pre-state societies ordered high-trust roles by age and accumulated standing; long-distance trading guilds ranked members by cohort and partner seniority; the magistracies of the Roman Republic ascended a legally codified *cursus honorum*; modern open-source platforms layer contributors by reviewer privilege, maintainership, and committer status. In each setting the rank order is publicly observable even when individual transactions are not, and trust is rationed in proportion to standing. We call such an arrangement a *hierarchical community* and, in this paper, we develop a theory of when it sustains cooperation, how large it can scale, and the conditions under which it persists and out-grows other more egalitarian alternatives.

We present a model capturing a basic common architecture we call a *hierarchical equilibrium with trade selectivity (HE-TS)* that is shared across numerous cases that differ in period, technology, and scale. Ranks are coarse and public; individual interactions are not; the flow of trade is concentrated at the higher rungs; migration to a less hierarchical, more egalitarian alternative is often feasible but still not chosen. The Roman Republic of the third through first centuries BCE is the central case we analyse: a quinquennial census producing a public record of ranks, a legally ordered *cursus honorum* of magistracies that allowed individuals to be entrusted with greater responsibilities, and a sequence of incorporations of new non-ethnically Roman members that scaled the citizen body from a few thousand to of the order of millions while preserving the hierarchical ladder.

We contrast these arrangements with Rome's major competitors, particularly Carthage, but also Athens, Larissa and other Greek city-states, which we characterize as sustaining cooperation through *identity investments*. These societies created culturally rich, internally more egalitarian, but ultimately ethnically-based polities with large barriers to the entry of new citizens and without such a well-defined hierarchical ladder. We show that such limits in their admission of new citizens also mean that these polities grew by orders of magnitude less, and it was ultimately the ability of the hierarchy to scale that gave the Roman Republic a manpower advantage that would win them supremacy over the Mediterranean. We further discuss how the evolutionary stability of hierarchies may explain why they emerge, persist and grow across history, from early societies to online communities.

We present two centrepiece results that organise the analysis. *Scalability* (Corollary 2) records that in the hierarchy with trade selectivity, the maximum community size compatible with cooperation grows linearly in the capacity of a senior agent, where capacity is the number of distinct partners that can be served per period. Our *dynamic evolutionary robustness* result (Proposition 6) further records that, on a parameter region we make explicit in Section 3, the hierarchical community is internally sustainable against two egalitarian alternatives: a reciprocity-based benchmark and a community based upon requiring identity investments by its members. However when it becomes possible for a single person to do multiple tasks, and thus to selectively concentrate trust, the candidate paths of both these alternatives unravel under a one-shot cheat-and-migrate deviation: i.e. cooperation breaks down in egalitarian communities as their members develop incentives to cheat and defect to the hierarchy instead.

Despite this ability to scale, the welfare comparison between hierarchical communities and those with identity investments is two-sided. The identity-investment arrangement delivers a strictly more equal distribution of within-community payoffs; the hierarchical community delivers a strictly larger sustainable size. We formalize how the welfare differences and how these two margins—equality and scale—trade off against one another.

In our model, a finite population is partitioned at every date into communities of common size N. Within a community, members are indexed by a publicly observable seniority $\ell \in [0, 1]$, sorted into pairs by a matching technology, and play a bilateral trust game: a principal chooses a scale of trade $\lambda \in [0, 1]$ and an agent, observing λ, chooses to cooperate or defect. Cooperation is jointly efficient but privately tempting to defect upon. We use the moral-hazard formulation of Shapiro and Stiglitz [1984]; which provides a useful benchmark formalization with which to study institutions that sustain cooperation [Greif, 2001]. Bilateral trade outcomes are private; the seniority record is public and evolves upward through a stochastic kernel satisfying first-order stochastic dominance. The combination of private outcomes and a public rank reconciles anonymity in interaction with a publicly observable ranking.

We show that one institutional parameter is crucial: the task *capacity* J, i.e. the number of tasks that any agent can perform for others per period. If $J = 1$, each agent can only be entrusted with at most one task and hierarchy affects only the within-pair scale of trust. However, once $J \geq 2$, this allows individuals to *choose* to focus their trades on particular individuals. We show that in the resulting *hierarchical equilibrium with* *trade selectivity (HE-TS)*, this capacity binds: only the top fraction $1/J$ of members is active in any given period, and hierarchy reallocates the flow of trade itself toward high-seniority members. It is this selective-trade regime (where $J \geq 2$) that the paper develops.

Our analysis is organised around four main results. Proposition 1 characterises the unit-capacity benchmark $J = 1$. Equation 17 displays the core regime $J \geq 2$ in closed form in the deterministic large-population limit: members above an activity threshold serve J principals per period, the stationary cross-sectional density of seniorities is a power law, and the value function has two branches joined at the threshold. Corollary 2 records scalability. The dynamic-robustness Proposition 6 records dynamic evolutionary robustness against the identity-investment alternative. We show however that both centrepiece properties require the possibility of trade-selectivity, i.e. $J \geq 2$.

Section 4 describes how the Roman Republic provides a useful extended case in which each primitive of the model has a direct institutional counterpart and each proposition has an observable implication. The quinquennial census produced a public record of each citizen's rank; the *cursus honorum*, codified by the *Lex Villia Annalis* of 180 BCE, fixed the order, age thresholds, and widening *imperium* of the ladder; the bounded number of annual office-holders at each rung implements the selectivity the model formalises in J. The citizen body scaled by orders of magnitude over the third through first centuries BCE through four layers of incorporation, each widening the ladder without flattening it. We compare the Roman institutional system with that of Carthage, Athens, and Larissa. These were culturally-rich and attractive communities in which internal equality is high but which were limited by identity-specific entry barriers. We document how they resisted the incorporation of new non-co-ethnic citizens and grew slower (Figure 4). These cross-sectional comparisons illustrate the equality-versus-scale trade-off predicted by Corollary 2 and Proposition 6. We use the contemporaneous moment of one of the gravest Roman military defeats, that by the Carthaginian general Hannibal at Cannae (216 BCE), to illustrate how, to Hannibal's own surprise, subordinate groups in Italy refused to declare independence and instead revealed their preference to remain part of the Roman hierarchical system.

Related literature. Our paper builds upon rich literatures that emphasize the fundamental importance of trust in economic development, social networks, historical political economy, and economics more generally [e.g. Arrow, 1974, McMillan, 2002]. In particular, we build upon an important literature on how repeated-games with either private monitoring or community-based third party enforcement can sustain trust [Kandori, 1992, Ellison, 1994, Greif, 1993, Dixit, 2003, Wolitzky, 2015, Ali and Miller, 2016, Olszewski and Safronov, 2018a, Clark et al., 2021]. Existing work also formalizes the idea that the earning of publicly recognizable symbols or tokens can enhance cooperation.[1] We share the private-monitoring premise but replace belief-based public records of past trade with the institution of a public hierarchy— a coarse, slowly evolving public rank whose accuracy is not profitable to misreport on the candidate path. In our main application, the Roman Republic, we compare to its main rivals, Carthage and the Greek city-states, which adopted conventions that we argue mimic equilibria that involve *cultural capital* [Klein and Leffler, 1981, Iannaccone, 1992, Fryer Jr., 2002], specific investments in markers of distinct group membership that can also sustain cooperation in both smaller groups [eg Kranton, 1996, Carmichael and Macleod, 1997, Ramey and Watson, 2001, Sobel, 2006] and broader communities [Board, 2008, Friedman and Resnick, 2001, Carvalho, 2013, 2016, Mailath et al., 2016b, Bramoulle and Goyal, 2016].[2]

We show that without the possibility of entrusting particular individuals with more tasks ($J = 1$), the hierarchical equilibrium recovers the same existence region as an identity-investment arrangement: in essence spending time in a subordinate role in a community itself acts as costly investment in that community's identity. However, when individuals can be entrusted with more than one task ($J > 1$), and thus trade selectivity is possible, these institutions diverge: we show that hierarchies can sustain cooperation for populations that scale linearly with the number of those tasks.

Similarly, in the literature on cooperation in social networks [eg Jackson, 2003, Bloch et al., 2008], the concept of hierarchy tends to be tied to network centrality that is based

[1]For example, Olszewski and Safronov [2018b] and Wolitzky [2015] consider equilibria with *chip* strategies, where individuals give tokens (or chips) to other agents that do not fink in trust games. These tokens are publicly visible and thus aid in community enforcement: individuals who deplete their tokens are (inefficiently) prevented from trading (for a spell), in order to sustain cooperation among those that possess tokens.

[2]For example, Carvalho [2013] and Carvalho [2016] consider *identity-based organizations*: clubs that have minimal strictness requirements based upon the identity investments or socialization of agents. Unlike our identity investment model, individuals who interact with outsiders can be socialized into losing part of their identity investment. Thus identity-based organizations that wish to sustain high degree of club goods provision among their members adopt strict minimal requirements on the identity investments of those seeking to join. Bramoulle and Goyal [2016] present a model of favoritism, where surplus is transferred from outside groups to inside groups. Similarly, Mailath et al. [2016a] examine conditions when 'buying locally' may be preferred. The intuition is that individuals may 'sacrifice' by overlooking a better outside product if there is some form of reciprocity– if they too gain as producers from others focusing trades locally. As in clubs, specific identity investments can also create value for participants in firms, including through improved corporate culture [Kreps, 1990], reduced costs of coordination [Hermalin, 2013] and through learning [Van den Steen, 2005].

upon past trades or ties that link specific individuals.[3] In our hierarchical equilibrium, the social hierarchies we construct are "impersonal" in the sense that the actual agents in the hierarchy can change but they inherit the incentives of their rank and thus fully efficient exchange can be sustained. Instead, akin to military custom or the ideal type of a Weberian bureaucracy, individuals in our hierarchical equilibria *salute the rank* of others even if they have never encountered them before.[4]

Our analysis also differs from much of the existing literature on trust in its focus on endogenous group formation, hierarchical structures, and the problems associated with increasing population size. In this way it links to important, but mostly parallel literatures looking at cultural transmission [e.g. Boyd and Richerson, 1994, Bisin and Verdier, 2011, Doepke and Zilibotti, 2013] and the origins of hierarchy and formation of state-like institutions [Bates et al., 2002, Bentzen et al., 2017, Besley and Persson, 2009, North et al., 2009, Bowles and Choi, 2013, Seabright, 2013, Dow and Reed, 2013, Boix and Rosenbluth, 2014, Mayshar et al., 2015, 2022, Michalopoulos and Papaioannou, 2020, Allen et al., 2023, Flückiger et al., 2024] as we describe below.

The remainder of the paper proceeds as follows. Section 2 sets up the model. Section 3 states the main propositions. Section 4 presents the mapping of the hierarchical equilibrium to the Roman Republic in comparison to its rivals: Carthage, Athens, and Larissa. Section 5 concludes. Appendix A contains the proofs; Appendix B treats the three-individual ($N = 3$) case; Appendix C discusses how our model may also shed light on important puzzles such as the emergence of hierarchy in early human societies, the development of hierarchy among modern ethnic groups such as the Hausa, and the sustenance of cooperation in anonymous online communities.

2 The Model

We study an infinitely repeated trust game played within and across communities with endogenous mobility. The primitives build on the moral-hazard formulation of Shapiro and Stiglitz [1984] and on community enforcement [Greif, 1993, Kandori, 1992]. At each

[3] A related network literature [Bramoulle and Goyal, 2016, Jackson et al., 2012] emphasises how network density substitutes for institutional rank; our framework is complementary in asking what a public rank delivers when network density is fixed and private.

[4] The key distinction between an agent's rank and their personal identity naturally has a long tradition in sociology. In particular, we build on and further work, at least as early as White [1970], on mobility in hierarchical organizations and "chains of vacancies" created by openings at higher ranks of hierarchies (see also Gibbons [2005]).

date a finite population is partitioned into communities; within a community, pairs are formed, play a bilateral trust game whose intensity is chosen by the principal, and then each individual chooses whether to remain or migrate. Every member carries an observable rank, called *seniority*, that evolves stochastically over time; bilateral trade outcomes are observed only by the pair.

Players, periods, and communities. Time is discrete, $t \in \{1, 2, \ldots\}$. A population of $P \in \mathbb{N}$ risk-neutral individuals, indexed by $i \in I := \{1, \ldots, P\}$, is partitioned at every date into M communities of common size $N := P/M$, with $N \geq 2$, N even, and $M, N \in \mathbb{N}$. Each individual survives between periods with independent probability $\delta \in (0, 1)$; deceased individuals are replaced by newborns, so that the population is stationary.

Stage game (trade). Within a community, individuals are sorted by the matching technology defined below into ordered pairs (p, q), with p the *principal* and q the *agent*; $i, j, k \in I$ denote generic individual indices. The pair plays a sequential trust game: the principal chooses a scale of trade $\lambda \in [0, 1]$; the agent observes λ and plays $a \in \{\texttt{cooperate}, \texttt{defect}\}$. Payoffs are

$$u^P(\lambda, a) \;=\; \begin{cases} \alpha\,\lambda\,w & \text{if } a = \texttt{cooperate}, \\ -\,c & \text{if } a = \texttt{defect}, \end{cases} \qquad u^A(\lambda, a) \;=\; \begin{cases} (1-\alpha)\,\lambda\,w & \text{if } a = \texttt{cooperate}, \\ \lambda\,s & \text{if } a = \texttt{defect}, \end{cases} \tag{1}$$

with $w > 0$, $s > w$, $c \geq 0$, and $w > s - c$. Cooperation is jointly efficient; defection is privately tempting at any $\lambda > 0$. The cost c is a fixed exposure cost incurred by a principal paired with a defecting agent; it is not proportional to λ.

We analyse the limiting case $\alpha \to 0$, in which the agent appropriates almost all of the within-period surplus when she cooperates; the goal is to simplify the analysis while still guaranteeing the principals have an incentive to select the largest incentive-compatible λ.

Seniority and the advancement process. Each community is endowed with an ordered set of *seniority levels* $L := [0, 1]$; at every date t each member has a seniority $\ell_{it} \in L$. Newcomers—newborns and movers alike—carry no seniority record from a previous community; for that one transition we treat every newcomer as if her previous-period seniority had been zero, so her period-of-entry seniority is drawn from $G(\cdot, 0)$.

Individual seniority evolves via a family of cumulative distribution functions $G : (0,1] \times [0,1] \to [0,1]$: for every $\ell \in [0,1]$ and every $\tilde{\ell} \in (0,1]$,

$$G(\tilde{\ell}, \ell) \;:=\; \Pr\bigl(\ell_{i,t+1} \le \tilde{\ell} \,\big|\, \ell_{it} = \ell,\; i \text{ survives and stays}\bigr). \tag{2}$$

We impose three properties on G.

Assumption 1 (Monotonicity / FOSD)**.** For every $\tilde{\ell} \in (0,1]$ and every $\ell, \ell' \in [0,1]$ with $\ell' > \ell$, $G(\tilde{\ell}, \ell') \le G(\tilde{\ell}, \ell)$.

Assumption 2 (Upward mobility)**.** For every $\ell \in [0,1)$, $G(\cdot, \ell)$ places probability one on $(\ell, 1]$ and has no atom at ℓ. The top rank $\ell = 1$ is absorbing: $G(\tilde{\ell}, 1) = 0$ for every $\tilde{\ell} < 1$ and $G(1,1) = 1$.

Assumption 3 (Smoothness)**.** G admits a density $g(\tilde{\ell}, \ell) := \partial_{\tilde{\ell}} G(\tilde{\ell}, \ell)$ continuous on the open upper triangle $\{(\tilde{\ell}, \ell) : 0 \le \ell < \tilde{\ell} < 1\}$.

Assumption 1 means higher-ranked individuals face a weakly better promotion prospect. Assumption 2 restricts movement to be upward. Assumption 3 is a regularity condition used to keep the continuation-payoff operator well behaved.

Matching within a community. An exogenous *capacity* $J \in \{1, \dots, N-1\}$, with $J \mid N$, caps the number of distinct principals for whom an individual may serve as agent in a given period. Let $Q := N/J$ denote the number of active agents needed to serve the N tasks in a community.

At the start of every period, inside every community, the public randomisation device first orders members by seniority, breaking ties randomly. Write $i_{[1]t}, \dots, i_{[N]t}$ for the resulting order from lowest to highest seniority. The active agents are the Q highest-seniority members, $A_t := \{i_{[N-Q+1]t}, \dots, i_{[N]t}\}$. A *matching* is a collection of ordered principal–agent pairs (p, q) such that every community member appears exactly once as a principal, no individual is matched with herself, each active agent appears as agent for exactly J principals, and inactive members do not appear as agents. The device randomises uniformly over such matchings. If $\nu_t(i)$ denotes the number of principals for whom i serves as agent at t, the matching satisfies

$$\ell_{it} > \ell_{jt} \implies \nu_t(i) \ge \nu_t(j). \tag{3}$$

Thus capacity is filled from the top of the hierarchy downward. For a realised seniority vector $\boldsymbol{\ell}_t$, the *activity threshold* is

$$\hat{\ell}_t \ := \ \min_{i \in A_t} \ell_{it}, \tag{4}$$

with ties at $\hat{\ell}_t$ resolved by the public ordering.

Mobility and renewal. At the end of each period, surviving individuals choose $m \in \texttt{stay}, \texttt{leave}$. Movers are reassigned to other communities, where they are treated as newcomers. Both movers and newborns enter with initial seniority drawn from $G(\cdot, 0)$. Both movers and newcomers pay a cost $m \geq 0$ for joining. We assume that a vacant spot is always available elsewhere, so that individuals can freely relocate. Remaining vacancies are filled by newborns.

Initial condition. At date $t = 1$, each community's members are treated as newcomers: $\boldsymbol{\ell}_1$ is drawn i.i.d. from $G(\cdot, 0)$. Let $F_t : [0, 1] \to [0, 1]$ denote the empirical cross-sectional distribution of seniorities at date t. Under Assumptions 2–3, F_t is atomless below the top rank; the absorbing top rank carries non-negative mass. The steady-state cross-section F_∞ is induced by δ, G, and the mobility decisions on the candidate path; it is not imposed as a primitive.

Information and monitoring. Monitoring is of two kinds. Within a pair, the realised (λ, a) is observed only by the pair's two members; bilateral observations are erased whenever either partner dies or moves, so trade outcomes are private and local. In parallel, each community is endowed with an *institution*—a public record of the seniority vector $\boldsymbol{\ell}_t := (\ell_{1t}, \ldots, \ell_{Nt})$ within the community, with identities attached, and of the realised matching π_t—that every community member observes at the start of each period.

Histories, strategies, equilibrium. Let τ_i^c denote the period in which i joined her current community c. The period-t *public history* available to i is

$$h_{it}^{\text{pub}} \ := \ \big(\boldsymbol{\ell}_{c,s}, \pi_{c,s}\big)_{s=\tau_i^c}^{t},$$

and is reset to the empty history whenever i moves. Her *private history* is

$$h_{it}^{\text{priv}} \ := \ \big(\{(j, r_j, \lambda_j, a_j)\}_{j \in \mathcal{M}_i(s)}\big)_{s=\tau_i^c}^{t-1},$$

where $\mathcal{M}_i(s)$ is the set of community members i was matched with at date s, $r_j \in \{P, A\}$ is i's role, and (λ_j, a_j) is the realised trade. The *personal history* is $h_{it} := (h_{it}^{\text{pub}}, h_{it}^{\text{priv}})$.

Let $\mathcal{H}$ denote the space of personal histories h_{it}, and let Ω_t denote the realisations of the public randomisation device at date t. A *behaviour strategy* for i is a triple of measurable maps

$$\begin{aligned} \sigma_i^P &: \mathcal{H} \to \Delta([0,1]), \\ \sigma_i^A &: \mathcal{H} \times [0,1] \to \Delta(\{\texttt{cooperate}, \texttt{defect}\}), \\ \sigma_i^M &: \mathcal{H} \times \Omega_t \to \Delta(\{\texttt{stay}, \texttt{leave}\}). \end{aligned} \tag{5}$$

Let $\sigma := (\sigma_i^P, \sigma_i^A, \sigma_i^M)_{i \in I}$.

Definition 1 (Solution concept)**.** A profile σ is a *sequential equilibrium* [Kreps and Wilson, 1982, Mailath and Samuelson, 2006] if at every information set the continuation strategy is sequentially rational given a KW-consistent belief system over the unobserved components of the state.

Definition 2 (Stationary profile)**.** A profile σ is *stationary* if, for every i and every t, σ_i^P, σ_i^A, and σ_i^M depend on h_{it} only through the current seniority vector $\boldsymbol{\ell}_{c(i)t}$, the current matching $\pi_{c(i)t}$, and pair-specific private observations from the current period, including, for σ_i^A, the realised λ within the stage game.

Throughout we restrict attention to pure, stationary profiles. We note that strategies conditioning on a publicly observable, one-time *identity-investment* entry payment $m \geq 0$ are admissible under Definition 2: such a payment enters the public record at the moment of entry and conditions σ_i^P in exactly the way the seniority record $\boldsymbol{\ell}_{c(i)t}$ does. The *identity-investment* alternative analysed in Section 3.4 is therefore a profile within the same solution concept and is comparable to the hierarchical equilibria characterised below.

Period timing. Each period proceeds in six stages, (0)–(5) (see Figure 1): (0) seniorities transition via G and the public record is refreshed; (1) a matching π_t is drawn uniformly from those satisfying (3); (2) each principal observes his agent's seniority and selects λ; (3) the agent observes λ and selects a; (4) stage-game payoffs are computed and privately observed by the pair; (5) each surviving individual selects m, movers are reassigned, and newborns fill any remaining openings.

All proofs are in the appendix.

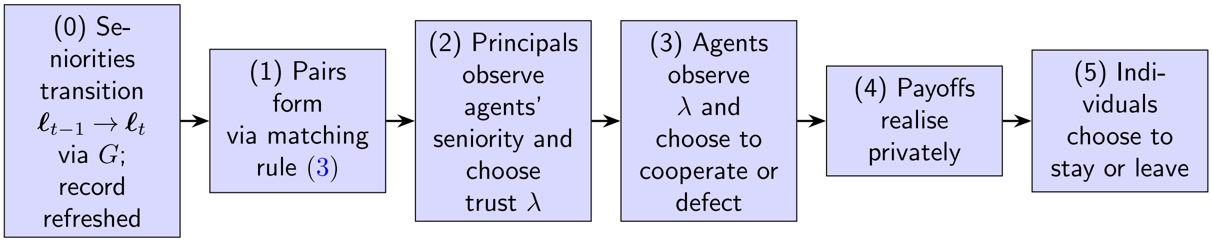


Figure 1: Timeline of a period, in six stages (0)–(5).

2.1 Discussion of modelling choices

When alternative formalisations yield the same qualitative conclusion, we adopt the one that makes our existence results harder.

A fixed c. The cost c is a fixed exposure cost from entering a trust relationship with a defecting agent, so a breakdown of productive trade imposes a fixed loss on the principal. Lower values of c make the within-group return constraint harder to satisfy.

Coexistence of public seniority and private trade. Bilateral trades are dyadic and observable in real time only to the pair; aggregating them into a community-wide record would require incentive-compatible reporting from every pair and is fragile [Kandori, 1992]. Seniority, by contrast, is a coarse, slowly evolving label tied to tenure. Because ranks pin down trust in the equilibria we study and cannot be profitably misreported, maintaining the public register is incentive-compatible on the candidate path.

Homogeneous payoffs and equal N. The stage game assigns a single (w, s, c) tuple to every pair; community size N is common. Unequal community sizes and bounded heterogeneity in (w, s, c) would not alter the qualitative conclusions below, with the role of the representative type played by the least-patient / most-tempted type on each margin.

Interpretation of the stage game. Cooperation here is not costly effort. Trade at scale λ generates joint surplus λw, of which the agent appropriates $(1 - \alpha)\lambda w$, approaching the entire surplus as $\alpha \downarrow 0$. The action `cooperate` is the agent's selection of the cooperative split rather than the larger one-period payoff λs obtained by defection; defection appropriates surplus rather than economising on effort. Two implications follow for the rest of the paper. First, being trusted at higher λ is a reward: it expands the agent's per-match payoff. Second, when capacity binds and only some agents are matched, being selected for trade is itself a benefit, not a burden. The hierarchy that emerges below should be read accordingly: higher seniority widens access to valuable trade opportunities.

Newcomer convention as a normalisation. Treating every newcomer as if her previous-period seniority had been zero is a normalisation. The analysis extends unchanged to any rank-reset kernel satisfying the same monotonicity and regularity properties, provided the binding incentive requirement remains at the top rank. However, it is true that partial seniority carry-over on migration would collapse the hierarchical advantage. Understanding those limits is left for future work.

3 Analysis

Within the pure, stationary sequential equilibria defined above, we study a family of profiles in which there are no entry fees ($m = 0$) and principals condition trust on the agent's seniority and agents condition cooperation and mobility on their continuation value in the community. The general hierarchical formulation at arbitrary capacity $J \geq 1$ admits two qualitatively distinct regimes. At $J = 1$ every member is active as an agent once per period, so hierarchy affects only the scale of trust λ conditional on seniority. At $J > 1$ only the top N/J members are active in any period.

Write $\varepsilon := (s - w)/w \geq 0$ for the one-period temptation normalised by the per-match surplus.

3.1 Maximal-trust hierarchical strategies

A *maximal-trust hierarchical strategy* specifies three on-path actions, as a function of the current seniority vector $\boldsymbol{\ell}_t$ and matching π_t:

(i) A principal paired with an agent at seniority ℓ chooses $\lambda = \ell$.

(ii) An agent of seniority ℓ chooses `cooperate`.

(iii) Every surviving individual chooses `stay`.

The label *maximal trust* reflects that $\lambda = \ell$ is the largest scale of trade the agent's incentive compatibility will allow: at any $\lambda > \ell$ the agent would strictly prefer to defect. As our punishment technology we use one-period Nash reversion inside the cheated pair. That is, the punishment phase triggered by a deviation to the above in $t-1$ or in t is a one period Nash reversion to no trust/defection in period t within the relevant parties. In the continuation game after any one stage deviation to the above, the principal always returns and the agent always leaves.

After any one-stage deviation, conditional on being re-matched to the same agent in the period after cheating, the cheated principal sets $\lambda = 0$ once, the defecting agent plays `defect` once. In the continuation game after any one stage deviation to the above, the principal always returns and the agent always leaves. Monitoring of the deviation is bilateral, so the reversion takes effect only if the cheated pair is re-matched at $t+1$. This is the weakest punishment that supports the construction; strengthening it can only enlarge the existence region.

Let $v(\ell)$ denote, on the candidate path, the discounted expected payoff of a surviving member at seniority ℓ immediately before the seniority-transition (stage (0)) of the next period. Because seniority cannot be lost on the candidate path and $\ell = 1$ is absorbing (Assumption 2), $v(1) = w/(1-\delta)$ when $J = 1$ and $v(1) = Jw/(1-\delta)$ when $J > 1$.[5] Away from the top rank, $v(\cdot)$ is pinned down by a Bellman functional depending on G and, under $J > 1$, on the stationary seniority distribution induced by G.

We refer to a maximal-trust hierarchical strategy that forms a stationary sequential equilibrium as a *hierarchical equilibrium* (HE) when $J = 1$ and a *hierarchical equilibrium with selective trade* (HE-TS) when $J > 1$. Two interpretive points are worth flagging. The trust scale $\lambda = \ell$ is a per-match opportunity, and the agent's share $(1-\alpha)\lambda w$ approaches the full surplus as $\alpha \downarrow 0$; reaching higher ℓ is therefore a reward, and active-set inclusion at $J \geq 2$ is selection into rents.

3.2 Lowest capacity: $J = 1$

At $J = 1$ every member is active once a period, so seniority acts only through the scale of trust $\lambda = \ell$. On the candidate path, the value function solves

$$v(\ell) \;=\; \int_{\ell}^{1} g(\tilde{\ell}, \ell)\big(\tilde{\ell} w + \delta\, v(\tilde{\ell})\big)\, d\tilde{\ell}, \qquad \ell \in [0,1), \tag{6}$$

[5] When $J > 1$, a member at the absorbing top rank is always active and, because her seniority is at the top, every principal matched with her extends full trust $\lambda = 1$. She serves J principals per period, each contributing w, so her per-period payoff is Jw.

with boundary $v(1) = w/(1-\delta)$.

With this notion at hand, the existence of a maximal-trust hierarchical equilibrium can be characterised through the joint satisfaction of two incentive constraints. They correspond to the two one-stage deviations available to an agent of seniority ℓ on the candidate path.

The first deviation is to *cheat and leave*: the agent plays `defect` at the current match, collects the one-period gain $\ell(s-w)$ relative to cooperation, and at the end of the period relocates to a fresh community where she re-enters as a newcomer at a draw from $G(\cdot, 0)$. Because seniority capital is not portable, relocation forfeits the discounted difference $\delta(v(\ell) - v(0))$ between the agent's continuation value at her current rank and the value of restarting. Ruling out this deviation gives the *migration constraint*, which we denote IC_L^H.

The second deviation is to *cheat and stay*: the agent plays `defect` but chooses `stay`. The one-period gain is again $\ell(s-w)$. The punishment is one-period Nash reversion within the cheated pair, which takes effect only if both members survive into $t+1$ (probability δ^2) and the same pair is re-matched (probability $1/(N-1)$ under random matching at $J = 1$). The associated expected loss has conditional mean $w \int_\ell^1 g(\tilde{\ell}, \ell)\tilde{\ell}\, d\tilde{\ell} +$ c: the agent's foregone per-match surplus tomorrow plus the exposure cost the cheated principal would otherwise have borne. Ruling out this deviation gives the *within-group return constraint*, which we denote IC_R^H.

The following proposition takes stock showing that these are the two key incentive constraints whose joint satisfaction characterises the existence of a maximal-trust hierarchical equilibrium.

Proposition 1. *Let $J = 1$. Under Assumptions 1–3, a maximal-trust hierarchical equilibrium exists if and only if there exists a pair $(g(\cdot), v(\cdot))$ such that v solves (6) with $v(1) = w/(1-\delta)$ and, for every $\ell \in [0, 1]$,*

$$\ell\,(s-w) \;\le\; \delta\big(v(\ell) - v(0)\big), \qquad \textit{with equality whenever } \ell < 1, \qquad (\mathrm{IC}_L^H)$$

$$\ell\,(s-w) \;\le\; \frac{\delta^2}{N-1}\Big(w \int_\ell^1 g(\tilde{\ell}, \ell)\tilde{\ell}\, d\tilde{\ell} + c\Big). \qquad (\mathrm{IC}_R^H)$$

The two constraints correspond to two different sources of discipline: seniority capital deters migration, while bilateral reciprocity deters opportunism within the community. The former constraint is the distinctively hierarchical one: the hierarchy makes migration costly without requiring an explicit entry payment. The maximal-trust requirement is reflected in the equality in IC_L^H for all $\ell < 1$: trust is expanded up to the point at which the agent is just willing to cooperate rather than cheat and leave. Economically the benchmark is analogous to any other institution that makes restarting costly. The latter constraint is the residual within-community enforcement constraint. It is the analogue of the standard bilateral reciprocity constraint: an agent who cheats but stays can be punished only if the same pair survives and is rematched next period.

What is new is that in a hierarchical environment this constraint is hardest to satisfy at the top of the ladder. The reason is simple. The current temptation to cheat, $\ell(s-w)$, is maximal at $l = 1$, while the punishment from a one-period loss of trust it is not. A top-ranked agent is already trusted at the maximal scale and has no further promotion rents to lose. Lower-ranked agents have to give up the additional trading surplus due to their future advancement. Thus the apparently counterintuitive fact that the top type is the tight type follows from a marginal comparison: senior agents have high continuation values in levels, but their one-period punishment is small relative to their current temptation.

Evaluating IC_R^H at $\ell = 1$ therefore gives the binding restriction on community size.

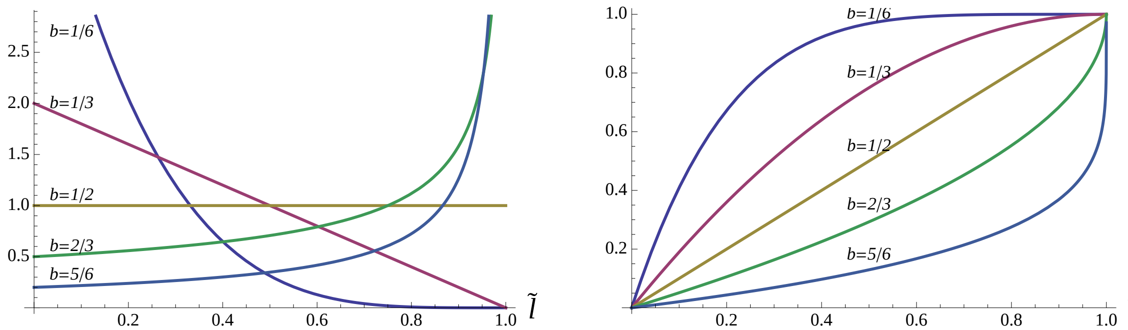
Corollary 1 (Bounded size)**.** *Under the conditions of Proposition 1, the maximum community size compatible with a maximal-trust hierarchical equilibrium at $J = 1$ is the constant*

$$N^*(1) \;=\; 1 \;+\; \frac{\delta^2(w+c)}{s-w},$$

independent of the advancement process g. In particular, at $J = 1$ hierarchy alone cannot scale: $N^*(1)$ is determined by the primitives (s, w, c, δ) and is invariant in G.

The bound $N^*(1)$ is invariant in the entire advancement process G: no choice of social mobility can relax it. Hierarchy at unit capacity is therefore a discipline mechanism that cannot, by itself, support cooperation at scale.

Social mobility, existence, and equilibrium structure. The characterization above expresses a hierarchical equilibrium as the solution to a system of functional equations, but does not, by itself, establish existence. To make further progress, we therefore focus on a tractable class of advancement processes that allows us to solve explicitly for the equilibrium objects. This serves two purposes. First, it provides a constructive existence result by directly exhibiting an equilibrium. Second, because the parametrization is sufficiently flexible to capture a wide range of advancement patterns,



(a) Density $g(\tilde{l}, 0)$ (b) Cumulative density $\int_0^{\tilde{l}} g(x,0)dx$

Figure 2: **Advancement Probabilities as Seniority Rises, by Social Mobility**: how the advancement probability and cumulative distribution change with current seniority l, for different levels of social mobility b. Panel (a) plots the density $g(\tilde{l}, 0)$ of next-period seniority $\tilde{l}$ given current seniority $l = 0$; panel (b) plots the corresponding cumulative distribution $\int_0^{\tilde{l}} g(x,0)\,dx$. For $b = 1/2$, the density is flat: advancement is independent of current seniority. For $b > 1/2$, there is *higher social mobility*, so newcomers advance faster and advancement slows as they become more senior. For $b < 1/2$, there is *limited social mobility*, so newcomers advance more slowly and advancement accelerates with seniority.

it allows us to isolate the role of social mobility and to understand more transparently how the hierarchical institution shapes incentives and outcomes.

To better understand how social mobility shapes incentives and outcomes, we focus on the following one-parameter family of advancement processes (or kernels), parameterized by $b \in (0, 1)$:

$$g(\tilde{l}, l) = \frac{1-b}{b} \times \frac{(1-\tilde{l})^{\frac{1}{b}-2}}{(1-l)^{\frac{1}{b}-1}}. \tag{7}$$

This parametrization allows us to capture *social mobility* within the community in a straightforward way. Let $\mu_{\tilde{l}|l}$ denote the expected seniority $\tilde{l}$, given current seniority l:

$$\mu_{\tilde{l}|l} = \int_l^1 g(\tilde{l}, l)\tilde{l}\; d\tilde{l}$$

Observe that:

$$\mu_{\tilde{l}|l} = l + b(1-l) \Leftrightarrow b = \frac{\mu_{\tilde{l}|l} - l}{1-l}. \tag{8}$$

Although imposing a functional form for g is obviously restrictive, the form still allows for a rich family of advancement processes. The parameter b can be interpreted as the proportion of the remaining gap an individual faces between current seniority and the highest possible level of seniority that the individual expects to cover in one period, conditional on surviving. Recalling that l is normalized so that it indicates the level of trust in the HE, it is thus a gauge of *social mobility* (see Figure 2). When $b \to 1$, all individuals expect to get to the top in one period; that is, there is extreme mobility, and individuals are trusted fully after their initial period. When $b \to 0$, for all $\tilde{l} > l$ we have $g(\tilde{l}, l) \to 0$; that is, there is no mobility. An intermediate case is $b = 1/2$. In this situation, the probability of advancement for an individual of seniority l is uniform between l and 1, and thus each period, each individual expects to cover half the distance between their current seniority and the top rank of $l = 1$.

Under this one-parameter kernel family it is possible to show that the incentive structure of the problem pins down b allowing us to simultaneously solve for a closed-form solution and to study the induced social dynamics (the formal proof is in Appendix A).

Proposition 2. *A hierarchical equilibrium exists if and only if*

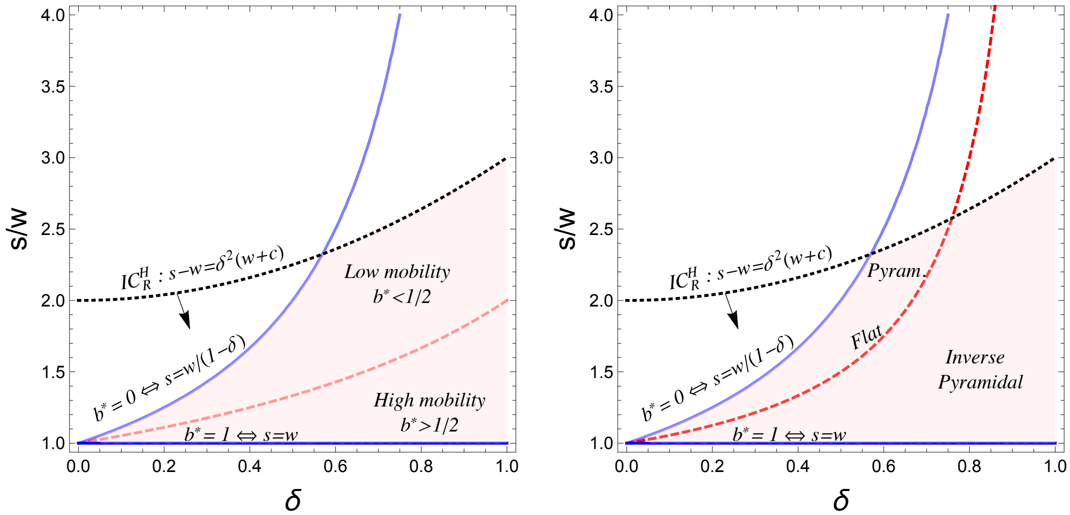
$$s < \frac{w}{1-\delta}, \qquad s - w \le \frac{\delta^2}{N-1}(w+c), \tag{9}$$

and is characterised by

$$v^*(\ell) = \frac{w}{1-\delta} - (1-\ell)\frac{s-w}{\delta}, \qquad b^* = 1 - \frac{s-w}{\delta s}. \tag{10}$$

Figure 3(a) depicts the existence region for a HE as a function of δ and s/w and relates them to the shape of the equilibrium advancement process. Recall that existence depends on four parameters, s/w, δ, N and c. In the figure, we set $c = 1$ and $N = 2$. The region enclosed by solid lines is defined by whether it is possible to find an advancement process (parameterized by b) to support a HE; thus, the upper and lower solid bounds represent the solution to (10) when $b^* = 0$ and $b^* = 1$, respectively. When s/w is such that s is close to the maximum long-term loss from deviating, $w/(1-\delta)$, then $b^* = 0$. On the other hand, when s/w is close to 1 (so that it is also possible to satisfy $s < w/(1-\delta)$), there is no incentive to defect and so cooperation can be sustained for all δ. Then, the HE has b^* close to 1 (in order to ensure that (IC_L^H) binds), so that newcomers advance to full trust very quickly.

The area below the dotted line is where the cheat and return constraint is satisfied. The dashed line in between solid lines in figure 3(a) is the locus where $b^* = 1/2$, that is, where the equilibrium is supported through a uniform equilibrium advancement process. It illustrates the boundary between *high (social) mobility* and *low (social) mobility* processes. Intuitively, the higher the incentive to defect, the higher the patience



(a) HE Existence region in the $(\delta, s/w)$ space assuming $c = 1$, $N = 2$ and shape of the equilibrium advancement process g^*.

(b) Same Existence region of subplot (a) but focusing on the shape of the induced steady state distribution f^*.

Figure 3: **Existence region of the hierarchical equilibrium ($J = 1$, $c = 1$, $N = 2$).** Panel (a) shows the existence region in the $(\delta, s/w)$ space, with the upper and lower solid lines tracing the solution of (10) for $b^* = 0$ and $b^* = 1$ respectively, and the dashed line marking the locus $b^* = 1/2$ (the boundary between high and low social mobility). Panel (b) shows the same existence region partitioned by the shape of the induced steady-state seniority distribution f^*.

required to sustain rapid advancement up the hierarchy.

We can also characterize the steady state density of the seniority level of individuals, denoted f. More precisely, $\int_{l'}^{l''} f(l)dl$ is the probability that an individual sampled at random at some arbitrarily large time t has seniority between l' and l''. This density is determined jointly by the survival rate δ and the advancement process g. Since seniority is normalized to correspond to the level of trust, the shape of f describes the distribution of trustworthiness in the community. f solves the following functional equation for $l \in (0, 1)$.

$$\overbrace{f(l)(1-\delta)}^{\#\text{ individuals dying}} + \underbrace{f(l)\delta\int_l^1 g(\tilde{l}, l)d\tilde{l}}_{\#\text{ ind. promoted to upper levels}} = \overbrace{\delta\int_0^l g(\tilde{l}, l)f(\tilde{l})d\tilde{l}}^{\#\text{ ind. promoted from lower levels}} + \underbrace{(1-\delta)g(0, l)}_{\#\text{ newcomers landing at level } l} \tag{11}$$

The above has exactly one solution given by:

$$f(l) = \frac{1-b}{b}(1-\delta)(1-l)^{\frac{1-b}{b}(1-\delta)-1}. \tag{12}$$

Computing f at the equilibrium $b^\star$ allows us to derive the equilibrium steady state distribution $f^\star$. Taking the derivative with respect to l rearranging, one can show that $f^*(l)$ decreases with seniority l if and only if:

$$\frac{s-w}{s} > \frac{\delta}{2-\delta}. \tag{13}$$

Observe that communities have *pyramidal* hierarchical structures—i.e., there are more individuals at the bottom than the top—when either the incentives to cheat are large $(s-w)$ or if patience is low. Conversely, higher social mobility and more *egalitarian* inverse-pyramidal communities are sustained in maximal trust hierarchies when the incentives to cheat and defect are weaker.

Figure 3(b) illustrates the existence region partitioned by the shape of the steady state distribution. Notice that the dashed line (representing a uniform distribution over seniority) would fall to the left of the dashed line if superimposed on Figure 3a (representing a uniform advancement process), and the wedge between them increases with δ. Intuitively, when individuals are long-lived, even with low social mobility, they will mass at the top of the distribution. So an equilibrium advancement process with low social mobility does not imply a pyramidal hierarchy in the steady state. However, a pyramidal steady state distribution does imply the presence of low social mobility.

3.3 The $J > 1$ regime

At $J > 1$, hierarchy affects not only the scale of trust but also the allocation of trade across agents. Recall from the matching technology that, for any realization of seniorities $\boldsymbol{\ell}$, the activity cutoff $\hat{\ell}(\boldsymbol{\ell})$ is defined as the lowest seniority among those agents who are assigned tasks, so that allocating J tasks to each agent with $\ell_i \geq \hat{\ell}(\boldsymbol{\ell})$ exhausts the total number of trades N. The key observation is that not all agents are active in a given period: only those above the cutoff engage in trade, while lower-ranked agents are idle. As a result, an agent's continuation value depends on her probability of being above the cutoff in future periods, which in turn depends on the entire realization of seniorities $\boldsymbol{\ell}$. Inclusion in the active set is selection into rents: an active agent serves J principals and captures the agent's share on each trade, so higher seniority widens access to these rents rather than enlarging an obligation.

In a maximal trade hierarchical the equilibrium payoff along the candidate path necessarily satisfies:

$$v(\ell_i, \boldsymbol{\ell}_{-i}) \;=\; \int_{\tilde{\ell}\in[0,1]^N} \Big(J\tilde{\ell}_i\, w\, \mathbf{1}\{\tilde{\ell}_i \ge \hat{\ell}(\tilde{\ell})\} \;+\; \delta\, v(\tilde{\ell}_i, \tilde{\ell}_{-i}) \Big)\, \mathcal{G}(\tilde{\ell}, \boldsymbol{\ell})\, d\tilde{\ell}, \tag{14}$$

for $\ell_i < 1$, with $v(1,\cdot) = Jw/(1-\delta)$ and $\mathcal{G}(\tilde{\ell}, \boldsymbol{\ell}) = \prod_j g(\tilde{\ell}_j, \ell_j)$. The equilibrium problem is inherently high-dimensional and hardly tractable. In particular, computing continuation values requires tracking the joint distribution of all agents' seniorities.

Before turning to the existence question, we characterise the existence of an HE-TS through the joint satisfaction of two incentive constraints, parallel to Section 3.2. They correspond to the same two one-stage deviations available to an agent on the candidate path; what changes under selective trade is the per-period gain from cheating and the rematch probability used in the punishment.

Consider first the *cheat-and-leave* deviation. An active agent of seniority ℓ_i serves J principals at the candidate scale of trust, earning $J\ell_i w$. If she defects and migrates, in equilibrium it is optimal to cheat simultaneously against all J principals: her one-period gain is therefore $J\ell_i(s-w)$. Relocation forfeits the discounted continuation $\delta(v(\ell_i, \boldsymbol{\ell}_{-i}) - v(0))$. Ruling out this deviation gives the *migration constraint*, denoted IC_L^{TS}.

The *cheat-and-stay* deviation is bilateral: defecting against one principal is unobserved by the other $J-1$ principals (per the information structure of Section 2), so each principal–agent relationship is checked separately. The one-period gain on a given match is $\ell(s-w)$. The punishment is one-period Nash reversion within the cheated pair, and takes effect only if both members survive (probability δ^2) and are re-matched at $t+1$. Under selective trade, the rematch probability for a top-ranked agent is $J/(N-1)$: the deviator at $\ell = 1$ is active with probability one, and each of her J principals draws her again with probability $1/(N-1)$ from the active set. As in Section 3.2, the within-group return constraint is tightest at the top of the hierarchy (we discuss why below). We refer to this constraint as IC_R^{TS}.

The next proposition formalises the characterisation; its logic parallels Proposition 1.

Proposition 3 (IC characterisation, core regime)**.** *Let $J \ge 2$. Under Assumptions 1–3, a maximal-trust hierarchical equilibrium with selective trade (HE-TS) exists if and only if there exists a tuple $(b, v(\cdot,\cdot))$ such that, for every seniority profile $\boldsymbol{\ell} \in [0,1]^N$, v satisfies the value functional (14) with $v(1,\cdot) = Jw/(1-\delta)$, and the following constraints* *hold:*

$$J\,\ell_i(s-w) \;\leq\; \delta\big(v(\ell_i,\boldsymbol{\ell}_{-i})-v(0)\big), \qquad \textit{with equality for } \hat{\ell}(\boldsymbol{\ell})<\ell_i<1, \tag{IC_L^{TS}}$$

$$s-w \;\leq\; \frac{\delta^2\,J}{N-1}\,(w+c). \tag{IC_R^{TS}}$$

Two features of the selective-trade environment require comment, however. First, the within-group return constraint is operative only for active types $\ell \in [\hat{\ell}(\boldsymbol{\ell}),1]$; below the cutoff the agent has no current match to defect on and the constraint is vacuous. Second, the rematch probability that enters the right-hand side of IC_R^{TS} is $J/(N-1)$.

Second, because the same factor J inflates both the on-path period payoff and the rematch probability, the slack profile in ℓ has the same monotonicity as in Section 3.2: the binding type is the one with maximal current temptation and zero scope for further continuation gain, namely $\ell = 1$. This is the genuinely novel mechanism relative to the unit-capacity case. At $J = 1$ the rematch probability is fixed at $1/(N-1)$ and hierarchy alone disciplines a deviator only through the seniority shortfall. At $J \geq 2$, selective trade raises the rematch probability for top-ranked agents by the same factor J that scales up their period payoff; the inflated probability is precisely what allows community size to grow with capacity.

Corollary 2 (Scalability)**.** *Under the conditions of Proposition 3, the maximum community size compatible with a maximal-trust HE-TS is*

$$N^*(J) \;=\; 1 \;+\; J\,\frac{\delta^2(w+c)}{s-w}, \tag{15}$$

Although the binding of IC_R^{TS} at the top is the formal parallel to Corollary 1, its economic content is different. Where Corollary 1 recorded that hierarchy alone caps community size at a constant of the parameters, the corollary that follows shows that under selective trade the maximum community size grows linearly in J and is unbounded as $N \to \infty$ and $J \to N$: hierarchy plus selective trade scales, where hierarchy alone does not. The comparative statics of $N^*(J)$ are invoked in Section 4 to trace the direction in which real-world confounders bias observed cross-sectional relationships between capacity and community size.

Capacity J, survival δ, per-match surplus w, and exposure cost c shift the ceiling upward; the temptation gap $s-w$ and the defect payoff s shift it downward. The signs are read off in Section 4 to sign the direction in which real-world confounders bias

Table 1: Comparative statics of the maximum community size.

Parameter θ	$\partial N^*/\partial\theta$	Sign
J (capacity)	$\delta^2(w+c)/(s-w)$	>0
δ (survival)	$2\delta J(w+c)/(s-w)$	>0
w (per-match surplus)	$J\delta^2(s+c)/(s-w)^2$	>0
c (exposure cost)	$J\delta^2/(s-w)$	>0
$s-w$ (temptation)	$-J\delta^2(w+c)/(s-w)^2$	<0
s (defect payoff)	$-J\delta^2(w+c)/(s-w)^2$	<0

observed cross-sectional relationships between capacity and community size.

Social mobility, existence, and equilibrium structure. To compute continuation values explicitly we exploit a large-population approximation. When N is large, the empirical distribution of seniority becomes approximately deterministic and converges to the steady-state distribution f. As a result, the cutoff $\hat{l}(\boldsymbol{\ell})$ stabilizes around a constant $\bar{l}$, defined by

$$J\int_{\bar{l}}^{1} f(x)\,dx = 1.$$

Intuitively, in large populations, agents behave as if the fraction of individuals above any given seniority level is fixed, so that the probability of being selected for trade depends only on one's own seniority. This observation allows us to approximate the continuation value by a function $v(l)$ that depends only on individual seniority, making the problem tractable. The next result formalizes this approximation. Importantly, it does not establish existence of hierarchical equilibria for finite N. Rather, it shows that, *if* such equilibria exist, their continuation values are well approximated by a tractable limit that can be characterized.

Proposition 4 (Large-N approximation of hierarchical equilibria)**.** *Fix $J \geq 2$ and parameters (s, w, δ, c) with $s > w$. Consider a sequence of economies indexed by $N \to \infty$. Suppose that, for each N, there exists a hierarchical equilibrium with trade selectivity, with continuation value*

$$v_N(\ell_i, \boldsymbol{\ell}_{-i}).$$

Assume that the empirical distribution of seniorities converges uniformly in probability to a deterministic distribution F with density f. Let $\bar{\ell}$ be the deterministic activity cutoff defined by

$$J\int_{\bar{\ell}}^{1} f(x)\,dx = 1.$$

Then there exists a deterministic continuation value $v(\ell)$ such that, for every fixed

$\ell_i \in [0,1]$,

$$v_N(\ell_i, \boldsymbol{\ell}_{-i}) \xrightarrow{p} v(\ell_i).$$

The limiting value function v is continuous and solves the deterministic cutoff problem associated with $\bar{\ell}$.

The proposition shows that, conditional on existence, the equilibrium can be approximated by a tractable one-dimensional problem in which continuation values depend only on individual seniority. The only source of approximation error is the randomness of the finite-N cutoff $\hat{l}(\boldsymbol{\ell})$. As N grows, the empirical distribution of seniorities stabilizes, so that $\hat{l}(\boldsymbol{\ell})$ converges to a deterministic cutoff $\bar{l}$, and the dependence of continuation values on the full state becomes negligible. Appendix A.7 provides numerical evidence on the magnitude of this approximation: simulating the stationary cross-section induced by the equilibrium advancement process, we show that the empirical cutoff is tightly concentrated around $\bar{l}$ even for moderately large populations. In particular, for N as small as 500–1,000, the dispersion of $\hat{l}(\boldsymbol{\ell})$ is already very small, and it continues to shrink rapidly at the parametric $1/\sqrt{N}$ rate as N increases up to 10,000.

This observation allows us to study a reduced problem in which the cutoff is treated as deterministic. Importantly, this step does not construct an equilibrium of the finite-N game. Rather, it provides a tractable characterization of the continuation values that any such equilibrium must approximate when populations are large. The resulting value function is therefore best interpreted as a limiting object that captures the relevant incentive structure, rather than as a standalone equilibrium solution.

Working with this reduced problem is useful for two reasons. First, it allows us to derive closed-form expressions and evaluate the incentive constraints explicitly, showing that they are satisfied on a non-degenerate parameter region with $(s-w)/w > 0$, so that the mechanism is not knife-edge. Second, it makes transparent the role of selective trade in reallocating surplus toward higher ranks and raising continuation values, which is the key force behind scalability and evolutionary robustness.

The limit value function $v(l)$ can therefore be obtained by replacing the state-dependent cutoff $\hat{l}(\boldsymbol{\ell})$ with the deterministic cutoff $\bar{l}$. In the limit, an agent earns flow payoff Jlw

whenever her seniority is above $\bar{l}$, and earns zero flow payoff otherwise. Thus $v(l)$ solves:

$$v(l) = \begin{cases} \int_l^{\bar{l}} g(\tilde{\ell}, l)\, \delta v(\tilde{\ell})\, d\tilde{\ell} + \int_{\bar{l}}^1 g(\tilde{\ell}, l) \big(J\tilde{l}\, w + \delta v(\tilde{\ell}) \big) d\tilde{\ell}, & 0 \le l < \bar{l}, \\ \int_l^1 g(\tilde{\ell}, l) \big(J\tilde{l}\, w + \delta v(\tilde{\ell}) \big) d\tilde{\ell}, & \bar{l} \le l < 1, \\ \dfrac{Jw}{1-\delta}, & l = 1. \end{cases}$$

The first branch applies to agents who are currently below the activity cutoff. They receive no current trade payoff unless next period's seniority draw moves them above $\bar{l}$. The second branch applies to agents already in the active region; conditional on remaining above the cutoff, they serve J principals and earn $J\tilde{l}w$ in the period. The boundary condition follows because the top rank is absorbing and an agent at $l = 1$ is always active. For the family

$$g(\tilde{\ell}, l) = \frac{1-b}{b} \frac{(1-\tilde{\ell})^{\frac{1}{b}-2}}{(1-l)^{\frac{1}{b}-1}}, \tag{16}$$

following the same strategy used to derive the closed-form expression for the $J = 1$ case, one can derive the corresponding closed-form solution:

$$v(l) = \begin{cases} \left(\dfrac{1-\bar{l}}{1-l} \right)^{\alpha} \left(J \dfrac{w}{1-\delta} - J(1-\bar{l}) \dfrac{s-w}{\delta} \right), & 0 \le l < \bar{l}, \\ J \dfrac{w}{1-\delta} - J(1-l) \dfrac{s-w}{\delta}, & \bar{l} \le l \le 1, \end{cases} \tag{17}$$

where

$$b^* = 1 - \frac{s-w}{\delta s}, \qquad \alpha = \frac{1-b^*}{b^*}(1-\delta).$$

The parameters b^* and α carry transparent comparative statics: b^* falls in ε (more mobility at higher temptation, so as to deliver enough continuation surplus), α rises in ε (the stationary density tilts toward the top), and $\bar{\ell}$ rises in ε for fixed J (the active region shrinks). The cutoff $\bar{\ell}$ also rises in J for fixed ε, since the active fraction $1/J$ shrinks with capacity.

Existence ($N = 3$). To further reassure on existence, Appendix B solves the smallest non-trivial finite case explicitly. For $N = 3$, the equilibrium can be constructed by brute force: continuation values can be computed directly and the two incentive constraints verified in closed form. This exercise delivers two useful messages. First, it shows that the hierarchical equilibrium with selective trade is not an artifact of the large-N approximation: it exists already in the smallest environment in which capacity binds. Second, it illustrates transparently the mechanism at work. When $J = 2$, selective trade raises the rematch probability for the top-ranked agent and relaxes the return constraint relative to $J = 1$, generating a non-empty parameter region in which equilibrium exists under trade selectivity but not under the unit-capacity benchmark. While the $N = 3$ construction is not intended as a general existence proof, it provides a concrete finite-N benchmark that complements the large-N characterization and confirms that the underlying mechanism operates away from the limit.

3.4 Identity-investment alternative

An natural alternative to hierarchies, one prominent in the literature on group-based cooperation [Carvalho, 2013, 2016, Akerlof and Kranton, 2000, Fryer Jr., 2002] and one that we describe practical analogues to below as well, sustains full trust through an *identity investment*, a sunk cost paid at entry rather than through accumulated seniority capital. We characterise that alternative in our own primitives, embed it inside the same game, and then study it in relationship with the hierarchy.

Notice that this institution is embedded as a special case of the model with $m > 0$ and a degenerate kernel in which all individuals transition to seniority $l = 1$ immediately after joining. In this case, the matching technology reduces to random assignment, subject to the same capacity constraint J: in each period, a subset of the population is selected at random to serve as agents, each carrying out up to J tasks, while the remaining individuals are inactive. Thus, identity investment differs from hierarchy only in the source of incentives, not in the underlying matching environment.

An *identity-investment strategy* sustains cooperation by replacing seniority capital with a single sunk cost paid at entry: newcomers pay a publicly observable investment on joining, are then fully trusted, cooperate, and stay. Formally, the strategy profile specifies (i) every newcomer joining a community pays a sunk cost $m > 0$ at entry; (ii) any principal matched with an agent chooses $\lambda = 1$; (iii) every agent plays `cooperate`; (iv) every surviving member chooses `stay`. The punishment phase is one-period bilateral Nash reversion inside the cheated pair, as in Section 3.1. We refer to a stationary sequential equilibrium in identity-investment strategies as an *identity-investment equi-*

librium (IIE). We restrict attention to $m \le w/(1-\delta)$ so that entry is individually rational.

Proposition 5. *An identity-investment equilibrium exists if and only if*

$$J(s-w) \;\le\; \delta\, m, \qquad (\mathrm{IC}_L^I)$$

$$s-w \;\le\; \frac{\delta^2}{N-1}(w+c). \qquad (\mathrm{IC}_R^I)$$

The minimum investment that supports cooperation is $\underline{m} = J(s-w)/\delta$*; the cheapest IIE is therefore characterised by the pair*

$$v_I^* = \frac{w}{1-\delta} - \underline{m} \;=\; \frac{w}{1-\delta} - J\frac{s-w}{\delta}, \qquad (18)$$

where v_I^* *is the entrant's discounted payoff* net *of the sunk investment.*

Notice that when $J=1$, the existence region of the hierarchical equilibrium of Proposition 1 coincides with that of the cheapest identity-investment equilibrium of Proposition 5: both admit the same maximum community size $N^*(1) = 1+\delta^2(w+c)/(s-w)$ and require $\varepsilon \le \varepsilon_0$. Moreover the entrant's net payoff and the established member's continuation value coincide across the two arrangements: $v^*(0) = v_I^* = w/(1-\delta) - (s-w)/\delta$ and $v^*(1) = w/(1-\delta)$.

3.5 The dynamic evolutionary robustness of hierarchies

We now compare a hierarchical community with selective trade to a community sustained by identity investments. The two environments share the same primitives and matching technology, but differ in how entry costs are generated. Identity-investment communities impose a sunk cost at entry. Hierarchies instead impose an implicit cost through low initial seniority and delayed access to trade.

As noted above, at $J = 1$ these two mechanisms are equivalent: the loss borne by a newcomer in the hierarchy exactly replicates the minimum investment required to sustain cooperation under identity investment, and the existence region is identical. The distinction emerges only when $J > 1$. Selective trade raises the value of reaching high ranks, thereby increasing the continuation value of entering the hierarchy.

Let $v(0, J)$ denote the continuation value of entering the hierarchical community as a newcomer from the general characterization (17) expressed in terms of primitives.

$$v(0,J) = \frac{w}{1-\delta} - \frac{s-w}{\delta}\, J^{-1/\alpha}, \qquad \alpha = \frac{(1-\delta)(s-w)}{\delta s - s + w}. \tag{19}$$

For comparison, the cheapest identity-investment equilibrium yields entrant value

$$\frac{w}{1-\delta} - \frac{s-w}{\delta}. \tag{20}$$

We say that a hierarchical equilibrium with selective trade is *evolutionarily stable* against the identity-investment alternative if two conditions hold. First, cooperation in the identity-investment community must be unstable to deviations toward the hierarchy. Second, the hierarchy must be attractive to entrants, so that such deviations are not reversed. Evolutionary stability requires two inequalities. The first condition ensures that an established member of the identity investment community matched to J principals finds profitable defection and migrating to the hierarchy, obtaining a current gain $J(s-w)$ and a continuation value $v(0,J)$:

$$J(s-w) > \delta\left(\frac{w}{1-\delta} - v(0,J).\right) \tag{ES-D}$$

The second condition ensures that the hierarchy is attractive at entry relative to the identity investment community with $m = (s-w)/\delta$[6]

$$v(0,J) > \frac{w}{1-\delta} - \frac{s-w}{\delta}. \tag{ES-A}$$

Condition (ES-D) guarantees that agents have an incentive to leave the identity-investment community, while Condition (ES-A) guarantees that the hierarchy absorbs these deviations: all incumbent members of the hierarchy prefer to remain in it, and newcomers prefer to enter it.

Proposition 6 (Evolutionary stability)**.** *The hierarchical equilibrium with selective trade is evolutionarily stable against the identity-investment alternative if and only if* $J > 1$.

At $J = 1$, the two conditions hold with equality. In this case, the hierarchy replicates

[6] Condition (ES-A) uses the entrant value from the cheapest identity-investment equilibrium at $J = 1$, with $m = \underline{m}(1)$. This is conservative: allowing m to rise with J would only strengthen the result. Holding m fixed makes clear that the effect is driven by selective-trade rents in the hierarchy, not by a mechanical increase in identity-investment costs.

exactly the incentives generated by identity investment: the implicit entry cost,

$$\frac{s-w}{\delta},$$

coincides with the minimum investment required to sustain cooperation. For $J > 1$, both conditions hold strictly. Substituting (19) into (ES-D) and (ES-A) yields

$$J > J^{-1/\alpha} \qquad \text{and} \qquad J^{-1/\alpha} < 1,$$

which are both satisfied if and only if $J > 1$.

The underlying mechanism is that the entrant value $v(0, J)$ is strictly increasing in J:

$$\frac{\partial v(0,J)}{\partial J} = \frac{s-w}{\delta}\frac{1}{\alpha}J^{-1/\alpha-1} > 0.$$

Higher capacity raises the payoff from reaching high ranks, since those ranks serve more principals. Although a larger J also makes access to the active set more selective, the increase in rents dominates. As a result, the implicit entry cost of the hierarchy declines from

$$\frac{s-w}{\delta} \quad \text{at } J = 1$$

to

$$\frac{s-w}{\delta}J^{-1/\alpha} \quad \text{at } J > 1,$$

which is strictly smaller.

The comparison should not be read as saying that hierarchy is unambiguously more efficient than identity investment. The point is rather that the two institutions dissipate surplus in different ways. Identity investment disciplines migration by burning resources at entry. Hierarchy disciplines migration by delaying access to scarce trading positions. When $J = 1$, this distinction is only cosmetic: the delayed access cost exactly replicates the minimum identity investment. When $J > 1$, however, selective trade changes the nature of the delay. The trades not allocated to newcomers are not destroyed; they are reallocated to senior agents who can serve several principals. The entrant therefore gives up current access to trade in exchange for an option to reach a position that commands multiple trades in the future. This option value lowers the effective entry cost of the hierarchy relative to the sunk investment required by the identity-investment benchmark. This is also why the result is not a general welfare ranking. Identity investment delivers full trust and a more equal allocation among those who have paid the entry cost, but it does so by imposing a sunk cost. Hierarchy avoids that sunk cost, but replaces it with unequal access to trade.

The advantages of hierarchy in this comparison comes from the fact that, under selective trade, hierarchies allow communities to scale, and prove to be robust even when the possibility of entrusting individual members with multiple tasks would lead to a breakdown of cooperation in the more internally egalitarian identity-investment communities. These two properties, of scalability and evolutionary robustness, we argue, are crucial for understanding why hierarchical communities often out-compete others, despite the attractive features of egalitarian societies. The rise of the Roman Republic to supremacy in the Mediterranean against such alternatives provides a useful, focal, example to which we turn to next.

4 Application: The Roman Republic

An important aspect of our model is that, unlike social networks which are often person-specific, social hierarchies are *impersonal* and the patterns of cooperation in a social hierarchy can survive beyond the individuals that populate them. In endogenous social network theory [eg Jackson et al., 2012], ties are based upon an individual's trading history. Instead, in our framework, centrality in the network of trust relationships derives from an individual's office or status. Agents *salute the rank* and can trust senior individuals even if they have never met before. Thus, the social structure can out-live an individual agent or hereditary lineage. Indeed, the development of institutions that support such impersonal relations has been seen as critical for the historic expansion of trade [Greif, 2001].

The impersonal hierarchy in our model is furthermore akin to a concept of citizenship which was also novel in the classical world, that of the Roman Republic (ca. 2nd-3rd centuries BC).[7] What allowed the Roman Republic to scale, granting ever-broadening citizenship, while its principal Mediterranean contemporaries and rivals—Carthage, Athens, Larissa, and the other Hellenistic poleis—did not? The Roman Republic provides a useful extended case in which every primitive of Section 2 has a direct institutional counterpart and every proposition of Section 3 has an observable implication. As Figure 4 shows, Rome was able to scale in its citizenship in a manner that was ultimately unmatched by its contemporaries, becoming the largest pre-modern community for which a codified, multi-tiered seniority ladder is documented. Further, as

[7] We are grateful to Ian Morris for pointing out these parallels with our model.

we describe below, the institutional facts—a quinquennial census producing a public rank record, a legally ordered sequence of magistracies (the *cursus honorum*), and successively wider grants of citizenship and office access—align tightly with the seniority record $\boldsymbol{\ell}_t$, the capacity J, and the scalability and dynamic-robustness predictions of the model.

We argue that an important contrast exists with Rome's contemporaries, like Carthage, Athens, Larissa and other Greek city-states, which were organized along traditional ethnic lines that mimic the identity investment equilibrium. Citizenship derived mainly from ethnic lineage. These polities developed rich and distinctive markers of their individual civic cultures, were internally more egalitarian, and therefore attractive to their members, but maintained exclusivity to outsiders that prevented scale. We argue that Rome's relative ability to scale in terms of people and thus manpower, on the reading developed below, is why it ultimately absorbed or out-lasted them in the Mediterranean.

In what follows, Section 4.1 maps the census and the *cursus* to model primitives and times the expansion of the ladder. Section 4.2 interprets Rome's incorporative expansion in light of the properties of the trade selectivity hierarchy and presents the Social War as a Part C stress test. Section 4.3 frames Carthage, Athens, and Larissa as identity-investment communities and discusses the relative properties of exclusivity and scale.

4.1 The census and the hierarchy of the Republic

> *Equality itself is inequitable, since it allows no distinction in rank.*
> - Cicero, *De Republica, I. 43*, 54 BCE

The public seniority record: the *census*. Consistent with the model, Rome had a public record of seniority in the form of a *census* conducted every five years (*lustrum*). Rather than merely a regular enumeration, which was common in other cities as well, the Roman term *census* itself had a different meaning, deriving from an Indo-European root for "siting a man ... in his ... *correct place in the hierarchy*, with all the practical consequences that this entails, and doing so by a *just public assessment*, by a solemn act of praise or blame '.[8]

[8] G. Dumezil, *Servius et la fortune*, 1942, p. 188 cited in Nicolet [1980], pg. 50, (italics ours).

Oversight of the census itself was entrusted to two *censores*, themselves chosen from among the most senior members of the hierarchy of the Republic— those who had already served as consuls.[9] Every five years, each citizen was required to appear, declare property under oath, and was assigned by the censores to a census rank whose voting weight, military obligations, and rank in the *comitia centuriata* (the property-class-based voting assembly) were publicly recorded. These regular public hierarchical census exercises were conducted both in Rome itself, and would be replicated in each city and colony of the Republic.[10]

Two pieces of the record correspond to the seniority label ℓ of Section 2: the census class, discretising the citizen body into five property-defined tiers plus the *capite censi* (citizens without property); and, for those close to the top of the hierarchy, membership with the 'Senate' (the term *senex* itself meaning 'elder'). Within this *album senatorium*, the censores further ordered each senator by the highest rank of public office - magistracy- each had held, which also established precedence. The class record and the *album* were both refreshed every five years.

Elite hierarchical ranks: *the cursus honorum* Further, the *Lex Villia Annalis* of 180 BCE codified a sequence of public offices with minimum ages and waiting periods: quaestor (at thirty), aedile or plebeian tribune (optional), praetor (at thirty-nine), consul (at forty-two), and censor.[11] Each rung conferred strictly wider *imperium* (the legal authority of a Roman magistrate). Access to high-stake trade—commands, state contracts, patronage networks—was concentrated at the top. The legally bounded number of annual office-holders at each rung (two consuls, eight praetors from Sulla, twenty quaestors) implements the capacity constraint of Section 2: not every senator is active in a given year, and activity is allocated by rank through annual election rather than by random draw.

Three mappings organise our application of the HE-TS to the Roman Republic. First, the census and the *album senatorium* are the empirical counterpart of $\boldsymbol{\ell}_t$: identities attach to ranks, ranks are observable at every *lustrum*, the record is refreshed by a designated authority, and bilateral patron–client transactions remain off the record. Second, the *cursus honorum* is a finite discretisation of $\ell \in [0, 1]$ with access to high-

[9]On the *professio* under oath, the public display of the *tabulae censoriae* (the official census tablets), and the quinquennial rhythm, see Nicolet [1980], pp. 49–88; on the censorial authority and the *lectio senatus* (the censorial revision of the senate roll), Lintott [1999], ch. 6.

[10]The *Tabula Heracleensis* (ca 80-70BCE) provides an inscription detailing the rules of conduct for all citizens of municipia and colonies just after the Social War (see below) and was explicitly modelled on those of Rome itself. See Nicolet [1980],pg 61-62.

[11]Lintott [1999], pp. 94–108; Brennan [2000], vol. II, pp. 387–421.

stake trade rationed at the higher rungs; widening of the ladder over time maps to increases in J. Third, as noted above, the censors were senior ex-consuls and the censorship was the terminal office, so those who revised the record had climbed it and had no further advancement at stake.

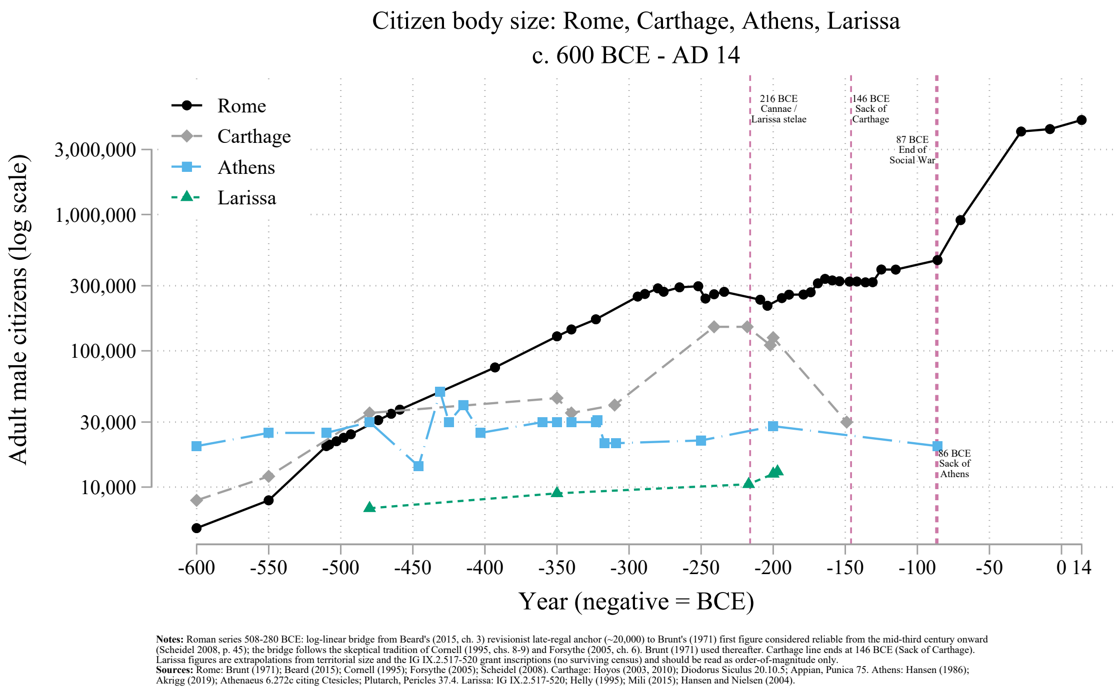


Figure 4: The Evolution of Citizen Populations in the Mediterranean

4.2 Scalability: the expansion of Roman citizenship

Corollary 2 says that N^* grows linearly in J with slope $\delta^2(w+c)/(s-w)$. Rome's growth from a city-state of a few thousand adult male citizens in the early Republic to roughly more than two hundred thousand at the outbreak of the Second Punic War and approximately six million at the universal grant of citizenship to all free persons in the Empire 212 CE is the empirical counterpart (Figure 4).[12] The Roman hierarchy expanded over five centuries, resolving potential and actual conflicts through the addition of rungs at the bottom and the relaxation of who could occupy them, never by the removal of intermediate offices. It is worth highlighting four expansions that are particularly noteworthy in this regard.

[12] Citizen counts from Brunt [1971], pp. 13–22 and p. 111, for the Republican figures; for the post-Italian-franchise population, pp. 89–115.

The *Conflict of the Orders* (c. 494–287 BCE). While initially only members of the oldest 'patrician' families were eligible for the highest rank offices, this early-Republican constitutional struggle between patricians and plebeians resulted in legislation that opened plebeian access to the top post of the consulship, with the *Lex Licinia Sextia* of 367 BCE reserving one of the two annual consulships for plebeians. Successful plebeian families thus were able to elevate the formal census status in the hierarchy.

The Latin incorporation (fourth century BCE) granted non-Romans from the region, the *Latini*, first citizenship without voting rights, but later full citizenship. These were not empty letter extensions: focal examples include provincials like Gaius Marius, who would command the Roman armies during the *Social War*, (see below) and Cicero himself.[13] While it is true that the elevation of such 'new men' was rare—the *nobilitas* retained a disproportionate share of consulships— Marius's significance for the model is that the path from the periphery of the citizen body to the apex was not legally closed—the advancement kernel G has positive support at the bottom of the seniority distribution.

The Social War, 91-87BCE By end of the second century BCE, Italian non-Roman citizens (*socii*) contributed the majority of the military manpower drawn on by Rome, bore comparable burdens to citizens, and received a share of booty and land at the discretion of Roman commanders, but they lacked voting rights and were excluded from the higher ranks of the hierarchy (the *cursus honorum*) because only citizens could hold Roman magistracies.[14]

The acquisition of new territories in the East also created new opportunities for those at the top of the Roman hierarchy (i.e. enlarging J). While political debates raged in Rome about extending citizenship, the Italian allies chose to set up their own *parallel* Republic. Crucially, their choice of structure was to form a confederation closely *replicating* the Roman hierarchical institutional form. The confederate capital at Corfinium was renamed *Italia*; the rebels issued coins bearing that name, constituted a Senate, elected dual consuls, and instituted a hierarchical ladder of offices.[15] Thus, they did not fight for an egalitarian alternative; they fought for replication of, and admission to, the hierarchical one. To resolve the ensuing *Social War*, Rome's ultimate response—the

[13]Cicero and Marius were both from provincials from the town of Arpinum. Born to a provincial non-senatorial Italian family and a *novus homo* (a senator with no senatorial ancestor), Marius would hold seven consulships between 107 and 86 BCE. Brunt [1971], pp. 89–115.

[14]Brunt [1971], pp. 89–115.

[15]Brunt [1971], pp. 89–115; Sherwin-White [1973], pp. 131–160.

Lex Iulia of 90 BCE and the *Lex Plautia Papiria* of 89 BCE—was to absorb the rival hierarchy, grant the Italians citizenship without flattening the ladder.[16] This, we argue, is consistent with the Part C mechanism at the population level.

Manumission By the late Republic, access to the hierarchy extended even to slaves: manumission rates were high enough that a non-trivial share of *liberti* (manumitted slaves) attained citizenship within a generation, and the children of *liberti* held full citizenship. Many slaves had a good chance of gaining freedom and full Roman citizenship in their lifetime, so much so that by the second century CE, the majority of citizens are believed to have had slave ancestry [Beard, 2015][pg.67].[17] This mobility extended to the top: the elite senatorial class itself became a multi-ethnic body [Beard, 2015][pg.68].[18] Remarkably compared to other contemporary states, but perhaps not surprisingly given its evolution over time, by 212 CE, the *Constitutio Antoniniana* granted citizenship to all freeborn residents of Roman territory. Yet, what is preserved across all these episodes is the hierarchical ladder: new cohorts enter at the bottom or at the quaestor level, never directly the top.

4.3 Carthage, Athens and the Greek poleis as identity-investment communities

In contrast, Rome's contemporaries, like Carthage, Athens, Larissa and other Greek city-states, were organized along traditional ethnic lines that, we argue, mimic the identity investment equilibrium. Citizenship derived mainly from ethnic lineage and city-specific cultural identity investments. Citizenship in Athens, for example, originally derived from membership in one of four tribes [Aristotle, 335BC][pg.56,65]. Citizens

[16] Sherwin-White [1973], pp. 150–174. New citizens entered the *cursus honorum* at the quaestorship and were distributed across tribes in ways that limited their immediate voting weight.

[17] An illustrative example is arguably provided by the famed Slave revolt, led by Spartacus (73–71 BCE). Spartacus' followers occupied a rank formally outside the citizen body; the revolt offered them an explicitly egalitarian outside option: the possibility of setting up new free communities beyond the reach of Rome. By 72 BCE the rebel army had defeated multiple Roman armies and reached the Alpine passes. (Plutarch, *Crassus*, 9–11; Appian, *Civil Wars*, 1.116–120.) Yet, despite this open route, the army turned south back into Italy, seeking to find an elevated rank within the Roman system rather than dispersing northward (Plutarch, *Crassus*, 10.1.). Had the rebels been pursuing an egalitarian alternative, the Alpine route was open; the southward turn is therefore arguably consistent with a continuation value within the Roman hierarchical structure that some perceived to have exceeded a potentially egalitarian outside option.

[18] Nicolet [1980], pp. 49–88, discusses the pathway and its demographic weight. The precise share is contested, but the qualitative point—that slave ancestry ceased to be a barrier to senatorial rank within two or three generations—is uncontroversial.

had specific privileges, many that built upon trust and specialization, that included relatively equitable economic opportunities and political rights, including in a number of prominent cases like Athens, *democracy* [Ober, 2015][chp 1]. In the presence of other competing city-states, there was arguably a truly remarkable efflorescence of cultural achievement in the classical period (consistent with a membership fee m that likely provided personal benefits to members rather than being purely a wasteful sunk cost).

However, nonetheless, there was systematic exclusion from these rights and opportunities of all outside the specific ethnic group of the city, with 'foreigners' including citizens of other city-states, ex-slaves and, in the case of colonies, even the original inhabitants. And, ultimately, these communities could not compete for scale with the Roman republic. Carthage with its far-flown Phoenician colonies and diasporas across the Mediterranean, from Gades (modern Cadiz) to the Levant, came among the closest, but even this ethnic group-delimited confederation ultimately could not continue to compete with military manpower afforded by the citizenry of the expanding Republic [Eckstein, 2008] (Figure 4).

Internal equality, external boundary. Classical Athens filled most magistracies by lot, paid jurors and assemblymen, distinguished sharply between *politai* and *metoikoi* (resident foreigners), and granted citizenship rarely. The internal life of the citizen body was, by ancient standards, egalitarian: Ober [2015], ch. 1, documents the relative flatness of political rights within the *demos*. Carthage was, in Aristotle's own comparison, organised along the lines of a Greek *polis*, with two annual *suffetes* (chief magistrates, elected for a one-year term) and a Council of Elders, but no codified sequence of lower offices and no legal pathway from subject to citizen.[19] The Phoenician core enjoyed a rough internal equality matching the Athenian pattern. In each case the costly common marker (descent, a shared cult, tribal membership) is the identity investment; the resulting community is internally trust-rich among those who bear it.

The identity-investment equilibrium has a number of attractive properties the hierarchical equilibrium does not: equality of per-period payoffs across active members; a low informational requirement, since no public seniority record is needed once the boundary is sunk; and low entry friction once inside. For polities the size of Athens or Phoenician Carthage, these were substantive welfare benefits.

Yet as described above, changes in technology or the expansion of trade which increases

[19] Aristotle [335BC], p. 56.

the capacity of individuals to be entrusted with additional tasks (a rising J) has a very different effect in identity-investment settings. Rather than increasing cooperative scale as in the formal hierarchies adopted by the Republic, to deter cheat-and-leave deviations by those who happen to receive more tasks requires the minimal membership fee m necessary to maintain cooperation to rise, *raising cultural exclusivity* and barriers to entry.

Exclusivity and Failure to Scale: the Larissa stelae, 215 BCE. The growing disparity between the rapidly scaling Roman Republic and the Greek city-states visible in Figure 4 was not lost on contemporaries, even while exclusivity was rising. These dynamics are well-reflected in a famous stelae from 215 BC that illuminates a contemporary Greek perspective on Roman institutions. It records an exchange to King Philip V of Macedon and the semi-independent Greek city of Larisa in Thessaly. In response to a previous letter suggesting that the Larisans should overcome their recent war-related depopulation by admitting new citizens from among its inhabitants, the Larisans had first admitted and then revoked the citizenship of more than 200 new inhabitants, the majority from Thessaly itself. In response, the King wrote:

> I hear that those who were granted citizenship in accordance with the letter I sent you and your decree, and whose names were inscribed (on the stele) have been erased. If this has happened, those who have advised you have ignored the interests of your city and my ruling. That it is much the best state of affairs for *as many as possible to enjoy citizen rights, the city to be strong and the land not to lie shamefully deserted, as at present*, I believe none of you would deny, and one may observe *others who grant citizenship in the same way. Among these are the Romans, who when they manumit their slaves admit them to the citizen body and grant them a share in the magistracies, and in this way have not only enlarged their country but have sent out colonies to nearly 70 places.*" [Austin, 2006][pg 157-159](italics ours).

In fact, the comparison between the "avaricious" Greeks and the "generous" Romans in their bestowal of citizenship has been credited both by contemporary observers and modern historians for Rome's relative ability to scale, and its ultimate dominance over the Hellenistic Mediterranean [Gauthier, 1974, Eckstein, 2008, Beard, 2015]

Hannibal's surprise at Cannae, 216 BCE. As we have discussed, the impersonal hierarchical institutional structure did not limit Rome by the size of any particular ethnic group and instead allowed Rome to accommodate a large, geographically dispersed and rapidly expanding population. That this expansion occurred in large part through consent rather simply coercion is visible following Rome's catastrophic defeat at Cannae in 216, at almost exactly the same time as the new citizens were being struck off the stelae by the Larissans. The Carthaginian general, Hannibal's, remarkable invasion of Italy was precipitated by a belief that by defeating the Romans in the field, he would allow other Italian cities to rise up against them. This he achieved at Cannae in 216 BCE—the worst single-day defeat in Roman military history, with between 55,000 and 70,000 Roman troops killed in a single day. Yet, despite this, and fatally for Hannibal's war plans, and ultimately Carthage itself, the majority of allied towns in Italy remained loyal to Rome [Eckstein, 2008][pg.20].

Hannibal's surprise at this loyalty— his strategy of releasing non-Roman Italian prisoners proved to be a costly contemporaneous indicator that he expected the opposite—is a valuable revealed-preference test in the historical record: the Italian elites faced a present option to regain their city-state autonomy, yet the majority chose to remain at their position on the Roman ladder. This revealed-preference is, we argue, the empirical counterpart of the formal argument in Section 4.2.[20]

Ultimately, our model suggests that the Carthaginians' and Greeks' unwillingness to admit members that could have been trustworthy and their attendant inability to scale relative to the Romans may have reflected the historic incentive problems they faced in maintaining cooperation given the presence of other proximate city-state communities in the Hellenistic Mediterranean. In contrast, the adoption by the Romans of a system of *impersonal* hierarchy with social mobility may have been a contributing factor in the Republic's rapid ascendancy and subsequent dominance.[21]

This is not to dispute that there was a trade-off across architectures that was substantive on both sides. The identity-investment equilibrium delivers higher per-period symmetry of payoffs across active members; the hierarchical equilibrium admits a strictly larger sustainable N^*. When the trust-relevant activities of Mediterranean political and eco-

[20]This interpretation also accords with that of Eckstein [2008], p. 20, who argues that allied loyalty was not merely an artefact of coercion but evidence of the value of remaining within the emerging Roman hierarchy.

[21]A traditional historical view credits Rome's military-oriented culture, including the glorification of warfare. As Eckstein [2008] points out, however, such cultures were pervasive among contemporary states, particularly amongst the successor-states of Alexander's empire, and military advances were rapidly shared and adopted across communities.

nomic life—long-distance trade, mass-mobilised legions, multi-front war—outgrew the size at which the identity-investment architecture could sustain them, the polity that scaled was the one that survived and absorbed its rivals. The historical record appears to selects the latter through Proposition 6.

To summarize, the key features of the model can be mapped to the case of the Roman Republic. The census and the *cursus honorum* are the public seniority record $\boldsymbol{\ell}_t$ and the capacity J; their self-enforcement through the censorship-from-*consulares* rule supports the consistency requirement on the candidate path. Rome's ability to scale is the reduced-form shadow of Corollary 2; the Social War test Proposition 6, Parts C. Carthage, Athens, and Larissa serve not as rival institution-free voids but as identity-investment communities— internally egalitarian, with attractive properties for their members, yet ultimately harder to scale. The dynamic channel of Part A arguably reflects what the two contemporaries, Hannibal saw at Cannae and King Philip V observed from Macedon.

5 Conclusion

Our comparative institutional analysis of the Roman Republic and the Hellenistic Mediterranean allows for a well-documented case for illustrating the relevant features of hierarchies in allowing the scaling of trust and cooperation in societies.

Further, we argue that the two centrepiece results we show, on both the *scalability* and the *dynamic evolutionary stability* of hierarchies once trading selectivity is possible, may explain why, as *all* individuals become more productive and can take on more tasks, whether due to changes in technologies or endowments, or the discovery of new trade routes, hierarchies may emerge to support cooperation at scale, and may even outcompete other existing and otherwise desirable institutional arrangements. In Appendix C, we discuss how our model may shed new light upon the birth of hierarchy in early human societies, even absent exogenous agent heterogeneity that might lead some to dominate others. We illustrate other features of the model through the example of institution-free egalitarian and hierarchical Hausa communities newly exposed to trans-Saharan trade. Finally, we discuss how hierarchies map upon the observable sustenance of cooperation among modern open source software communities as well.

It is worth noting that, unlike social networks which are often person-specific, the social hierarchies we have characterized are "impersonal". Individuals can trade with and trust others that they have never met, with rank providing sufficient information for trust, and the patterns of cooperation in a social hierarchy can survive beyond the individuals that populate them. Rank indicates that individuals benefit from remaining in a community, and thus are more trustworthy. Ultimately, the resilience of seniority-based cooperative hierarchies appears to mimic a number of environments where new venues for exchange have emerged. But it also hints at the dynamics of spontaneous order to expect in both historical settings and new markets where alternative institutions do not yet exist. In the absence of third-party enforcement, hierarchies are likely to emerge as a common early organizational form in new venues for trade.[22]

A Proofs

A.1 Preliminaries

We collect facts used below.

(P1) One-stage deviation principle. The infinite-horizon game has bounded stage payoffs and discount factor $\delta < 1$, and is therefore continuous at infinity. By Fudenberg and Tirole [1991, Theorem 4.2], a pure stationary profile is a sequential equilibrium if and only if no player has a profitable one-stage deviation at any information set consistent with a KW-consistent belief system [Kreps and Wilson, 1982]. Throughout, we verify the candidate profile by checking one-stage deviations at every on-path node and at the single off-path node reached after a one-stage deviation has occurred. After any unilateral deviation, the rest of the population continues to follow the candidate stationary profile σ^*, because deviations inside a pair are private (Section 2, "Information and monitoring") and all non-deviating members face the same on-path belief system

[22] Several directions remain open. Endogenising the investment in public rank structure—so that J is chosen by the community rather than taken as a primitive—would extend the scalability comparative static into a question about optimal institutional design; work in the community-enforcement lineage [Wolitzky, 2015, Olszewski and Safronov, 2018a, Clark et al., 2021] suggests that the mechanics would combine the incentive geometry of this paper with a set of participation constraints at the community level. Closed-membership extensions, in which a receiving community has the right of refusal, and partial seniority carry-over on migration, in which a mover retains a fraction of her accumulated rank, would alter the dynamic-evolutionary-robustness comparison in opposite directions and are natural targets for follow-up work. Finally, the empirical content of the comparative statics in Table 1 and the identification framework of Section 4 suggest a research programme in which the structural parameters (w, s, c, δ) are recovered from observable features of ladder-organised institutions across a range of settings from pre-industrial polities to contemporary platforms.

consistent with observing only the public seniority record. The one-stage deviation principle therefore applies [Fudenberg and Tirole, 1991, Lemma 4.3], and multi-stage deviations need not be checked separately.

(P2) One-period punishment value. Consider a principal–agent pair (p, q) that has just played a one-stage deviation at date t. By the punishment technology of Section 3.1, if both members survive to $t + 1$ and remain in the community, and if they are rematched to each other, the principal sets $\lambda = 0$ and the agent plays `defect`, yielding stage payoffs $(-c, 0)$ instead of the on-path $(\alpha \ell w, (1-\alpha)\ell w)$, so that the agent forgoes $w\ell$ and the principal incurs the exposure cost c. In the $\alpha \downarrow 0$ limit the entire within-period surplus is foregone by the agent; the punishment also saves the principal the exposure cost c she would have borne had the agent defected once more. The (unconditional) per-period loss to the agent from the punishment, evaluated under the on-path measure, is therefore $w\ell_{q,t+1} + c$ scaled by the joint survival probability δ^2 and the rematch probability. Under random matching ($J = 1$) that probability is $1/(N-1)$; under selective trade at $\ell = 1$ the top-ranked agent is active and every principal in her active set draws her with probability $J/(N - 1)$ by (P4).

(P3) Value at the top rank. By Assumption 2, $\ell = 1$ is absorbing: $G(\tilde{\ell}, 1) = 0$ for $\tilde{\ell} < 1$ and $G(1, 1) = 1$. A surviving member at $\ell = 1$ therefore remains at $\ell = 1$ forever. Under $J = 1$ she is active once per period and trusted at $\lambda = 1$, earning w each period; hence $v(1) = w/(1 - \delta)$. Under $J \geq 2$ she is active, trusted at $\lambda = 1$ by every principal who draws her, and serves J principals per period, earning Jw; hence $v(1) = Jw/(1 - \delta)$.

(P4) Independence of rematch across principals. At date $t + 1$, conditional on joint survival of the deviator and each of the J principals she cheated at date t (event of probability δ^2 for each principal, independently by the per-individual survival primitive of Section 2), the matching technology draws a matching π_{t+1} uniformly at random from those satisfying (3). Under the uniform draw, for each of the J principals the marginal probability of being matched again to the deviator is $1/(N - 1)$, and the events $\{\text{principal } p_k \text{ is rematched to the deviator}\}_{k=1}^{J}$ are conditionally independent given the active-set composition, because the uniform matching draws each principal's agent position independently subject only to feasibility

((3)) and because the deviator occupies at most one position per principal.[23] The expected joint loss from punishment, summed across the J principals, is therefore $J \cdot \delta^2(w \cdot [\text{conditional expected period payoff}] + c)/(N-1)$, used in Step 7 of A.2 at $\ell = 1$ and in A.1 Step 2(b).

(P5) Note on reversion to grim trigger. Let $\sigma^\star$ be a maximal-trust hierarchical sequential equilibrium under one-period Nash reversion. Replace the punishment phase within any cheated pair (p, q) by permanent bilateral reversion to $(\lambda, a) = (0, \texttt{defect})$, and retain all other strategy components. On the candidate path $\sigma^\star$ never invokes the punishment phase, so the on-path payoff is unchanged. Off the candidate path, the present value of the punishment loss strictly increases from $\delta^2(w \cdot [\text{period payoff}] + c)/(N-1)$ (or its J-scaled analogue under selective trade) to a strictly larger discounted sum that includes every future rematch of the cheated pair, not only the first one. The right-hand sides of the incentive constraints (IC_L^H), (IC_R^H), (IC_L^{TS}) and (IC_R^{TS}) therefore all weakly increase, so the existence region weakly enlarges; strictness obtains whenever the original right-hand side was attained with equality at the binding type. The same argument applies to grim trigger reversion under the HE-TS candidate.

A.2 Proof of Proposition 1

Set $J = 1$. On the candidate path every member is active exactly once per period and is trusted at $\lambda = \ell$, so (6) characterises the value function. We prove necessity, then sufficiency, then derive the closed form under (8).

Step 1 (Existence and uniqueness of v). Let $\mathcal{B} := C([0,1];\mathbb{R})$ equipped with the sup norm; $\mathcal{B}$ is a Banach space. Define $T : \mathcal{B} \to \mathcal{B}$ by

$$(Tu)(\ell) \;:=\; \int_\ell^1 g(\tilde{\ell}, \ell)\big(\tilde{\ell}w + \delta u(\tilde{\ell})\big)\, d\tilde{\ell}, \quad \ell \in [0,1), \qquad (Tu)(1) \;:=\; w/(1-\delta).$$

Under Assumption 3, g is continuous on the open upper triangle and (6) shows T maps bounded functions into bounded functions. T is monotone ($u \le u'$ pointwise implies $Tu \le Tu'$) and discounts by δ: $T(u+k) = Tu + \delta k$ for constant k. Blackwell's sufficient conditions [Stokey et al., 1989, Thm 3.3] hold, so T is a contraction on $(\mathcal{B}, \|\cdot\|_\infty)$ with modulus δ. A fortiori it has a unique fixed point v which is bounded; by the fundamental

[23] Formally: given the active-set realisation, the uniform measure over matchings factors as a product across principals up to a lower-order correction of size $O(1/N)$, vanishing in the mean-field limit $N \to \infty$; see Hildenbrand [1974, Ch. 2] for the analogous factorisation in large exchange economies.

theorem of calculus, $\tilde{\ell} \mapsto g(\tilde{\ell}, \ell)(\tilde{\ell} w + \delta v(\tilde{\ell}))$ is continuous on $(\ell, 1)$ under Assumption 3, so $(Tv)(\cdot)$ is differentiable on $[0, 1)$ with

$$v'(\ell) \;=\; \int_{\ell}^{1} \partial_{\ell} g(\tilde{\ell}, \ell)\big(\tilde{\ell} w + \delta v(\tilde{\ell})\big)\, d\tilde{\ell} \;-\; g(\ell, \ell)\big(\ell w + \delta v(\ell)\big). \tag{A.1}$$

Step 2 (Necessity of IC_L^H *and* IC_R^H*).* Fix $\ell \in [0, 1]$. On the candidate path an agent at ℓ earns $\ell w + \delta v(\ell)$ (period payoff plus continuation). Two one-stage deviations must be ruled out.

(a) Defect-and-migrate. The agent plays `defect`, collects ℓs, and at the end of the period chooses $m =$ `leave`. Her first-period seniority in the new community is drawn from $G(\cdot, 0)$; her continuation from the deviation onwards is therefore $v(0)$ as computed on the same candidate profile (the previous-period seniority of a newcomer is treated as zero). The deviation is not profitable iff

$$\ell s + \delta v(0) \;\leq\; \ell w + \delta v(\ell),$$

i.e. $\ell(s - w) \leq \delta(v(\ell) - v(0))$, which is IC_L^H (strict inequality for $\ell = 0$ holds automatically since the LHS is zero and the RHS is non-negative). At $\ell < 1$, strict slack in IC_L^H would permit the principal to extend trust at a scale $\lambda' > \ell$; the agent, facing instead $\lambda'(s - w) \leq \delta(v(\ell) - v(0))$, would still prefer to cooperate, which contradicts the maximality of $\lambda = \ell$ on the candidate path. Equality on $[0, 1)$ is therefore the maximal-trust requirement. At $\ell = 1$ maximality is automatic by (P3).

(b) Defect-and-stay. The agent plays `defect` and chooses `stay`. She collects ℓs today. At $t + 1$, with probability δ^2 both partners survive; under random matching ($J = 1$) the probability they are matched to each other again is $1/(N - 1)$. In that event the period payoff at $t + 1$ is 0 instead of $\tilde{\ell} w$ (for a drawn $\tilde{\ell}$) by the punishment technology (P2), a loss with conditional mean $w \int_{\ell}^{1} g(\tilde{\ell}, \ell)\tilde{\ell}\, d\tilde{\ell}$ for the agent plus the exposure cost c that the cheated principal would otherwise have borne.[24] Hence the deviation is not profitable iff

$$\ell s \;\leq\; \ell w \;+\; \delta^2 \tfrac{1}{N-1}\left(w \int_{\ell}^{1} g(\tilde{\ell}, \ell)\tilde{\ell}\, d\tilde{\ell} + c\right),$$

i.e. IC_R^H.

[24] The exposure cost c is borne by a principal paired with a defecting agent; on the candidate path the defector expects to face a cheated pair at $t + 1$ with conditional probability $1/(N - 1)$, and hence loses the per-period surplus $w\tilde{\ell}$ plus c relative to the on-path stream.

By (P1), these are the only one-stage deviations from the candidate profile.[25]

Step 3 (Sufficiency). Suppose v solves (6) with boundary $v(1) = w/(1-\delta)$ and (IC_L^H)–(IC_R^H) hold for every $\ell \in [0,1]$. By Step 2 no agent has a profitable one-stage deviation on the candidate path. Principals have no profitable on-path deviation: at any $\lambda < \ell$ the principal's payoff $\alpha\lambda w \to 0$ as $\alpha \downarrow 0$, so the indifference at $\alpha = 0$ is broken in favour of the largest IC-compatible $\lambda = \ell$. The mobility choice $m = \texttt{stay}$ is optimal because staying delivers $v(\ell) \geq v(0)$.

It remains to check the off-path continuation after a one-stage deviation in which an agent defects. The candidate continuation prescribes that the cheated principal returns, while the deviating agent leaves. The principal's return is optimal given that the agent leaves: the principal loses no future surplus from returning rather than relocating, and by returning remains available to impose the bilateral punishment should the deviating agent also return.

Now consider the deviating agent after the current-period cheating gain has already been obtained. Let $V^R_{\text{after}}(\ell)$ denote the agent's continuation value from returning to the same community after cheating, given that the cheated principal returns and the one-period bilateral punishment is applied if the pair is rematched. The cheat-and-return constraint can be written as

$$\ell(s-w) \;\leq\; \delta\big(v(\ell) - V^R_{\text{after}}(\ell)\big).$$

For $\ell < 1$, maximal trust implies that the cheat-and-leave constraint binds:

$$\ell(s-w) \;=\; \delta\big(v(\ell) - v(0)\big).$$

Combining the two inequalities gives

$$\delta\big(v(\ell) - v(0)\big) \;\leq\; \delta\big(v(\ell) - V^R_{\text{after}}(\ell)\big),$$

and hence

$$V^R_{\text{after}}(\ell) \leq v(0).$$

Thus, once the current cheating gain is sunk, the deviating agent weakly prefers leaving and restarting elsewhere to returning to the community with a cheated principal present.

[25] A defect-and-stay-then-leave-next-period deviation is equivalent to a defect-and-migrate deviation in which relocation is delayed by one period, and is strictly dominated by defect-and-migrate under $\delta < 1$.

If the inequality is strict, leaving is strictly optimal; if it binds, the candidate strategy selects leaving.

For $\ell = 1$, the same conclusion follows by continuity of v and the closed-form construction below, which implies

$$s - w = \delta\big(v(1) - v(0)\big).$$

Therefore the prescribed off-path continuation—principal returns, agent leaves—is sequentially rational. The prescribed one-period punishment itself is also sequentially rational: if the cheated pair is rematched, the principal sets $\lambda = 0$ and the agent plays `defect`, which is the stage-game Nash outcome in that bilateral relationship. By the one-stage deviation principle, these checks close the sufficiency argument.

A.3 Proof of Proposition 2

Under (8), substitute into (6) and differentiate:

$$v(\ell) \;=\; \tfrac{1-b}{b(1-\ell)^{(1/b)-1}} \int_{\ell}^{1} (1-\tilde{\ell})^{(1/b)-2}\big(\tilde{\ell}w + \delta v(\tilde{\ell})\big)\, d\tilde{\ell}.$$

Differentiating in ℓ and simplifying yields the first-order linear ODE

$$(1-\ell)\, v'(\ell) \;-\; \tfrac{1-b}{b}(1-\delta)\, v(\ell) \;=\; -\tfrac{1-b}{b}\,\ell w. \tag{A.2}$$

The homogeneous solution is $C_1(1-\ell)^{-\alpha}$ with $\alpha := \frac{1-b}{b}(1-\delta)$; a particular solution can be taken affine in ℓ: $v_p(\ell) = A + B\ell$. Plugging $v_p = A + B\ell$ into (A.2) gives $B - \alpha A = 0$ and $-B - \alpha B = -\frac{1-b}{b}w$, so $B = \frac{(1-b)w}{b(1+\alpha)} = w\,\alpha/[(1+\alpha)(1-\delta)]$ and $A = B/\alpha = w/[(1+\alpha)(1-\delta)]$. Hence

$$v(\ell) \;=\; C_1(1-\ell)^{-\alpha} \;+\; \frac{w}{(1-\delta)(1+\alpha)}\big(1+\alpha\ell\big).$$

The boundary condition $v(1) = w/(1-\delta)$ forces $C_1 = 0$ (the homogeneous solution blows up at $\ell = 1$ for $\alpha > 0$), so

$$v(\ell) \;=\; \frac{w(1+\alpha\ell)}{(1-\delta)(1+\alpha)} \;=\; \frac{w}{1-\delta}\left(\frac{1 + \frac{b}{1-b}(1-\delta)\ell}{1 + \frac{b}{1-b}(1-\delta)}\right). \tag{A.3}$$

Step 5 (Determination of b^*). Substitute (A.3) into IC_L^H at equality. Using $v(\ell) - v(0) =$

$\alpha\ell\, w/[(1-\delta)(1+\alpha)]$, equality in IC_L^H reads

$$(s-w) \;=\; \frac{\delta\,\alpha\,w}{(1-\delta)(1+\alpha)}. \tag{A.4}$$

Solving for α gives $\alpha^* = (s-w)(1-\delta)/[\delta w-(s-w)(1-\delta)]$; combining with $\alpha^* = \frac{1-b^*}{b^*}(1-\delta)$ yields

$$b^* \;=\; 1-\frac{s-w}{\delta s}, \tag{A.5}$$

matching the statement in (10). Note $b^* \in (0,1)$ iff $0 < s-w < \delta s$, i.e. iff $s < w/(1-\delta)$; this is the first inequality in (9).

Step 5a (Closed form of v^*). By (A.4), the quantity $\alpha^* w/[(1-\delta)(1+\alpha^*)] = (s-w)/\delta$. Substituting into (A.3):

$$v^*(\ell) \;=\; \frac{w(1+\alpha^*\ell)}{(1-\delta)(1+\alpha^*)} \;=\; \frac{w}{(1-\delta)(1+\alpha^*)} + \ell\,\frac{s-w}{\delta}.$$

Using $v^*(1) = w/(1-\delta)$ to solve for the constant, $w/[(1-\delta)(1+\alpha^*)] = w/(1-\delta) - (s-w)/\delta$, hence

$$v^*(\ell) \;=\; \frac{w}{1-\delta} \;-\; (1-\ell)\,\frac{s-w}{\delta}, \tag{A.6}$$

matching the closed form reported in (10) and the boundary value $v^*(1) = w/(1-\delta)$.

Step 6 (Binding of IC_R^H at $\ell = 1$). IC_R^H at generic ℓ reads

$$\ell(s-w) \;\le\; \frac{\delta^2}{N-1}\big(w(\ell+b(1-\ell))+c\big).$$

The RHS at $b = b^* > 0$ and fixed ℓ is linear in ℓ with slope $w(1-b^*) < w$; the LHS is linear in ℓ with slope $s-w$. The gap RHS$-$LHS therefore varies linearly in ℓ between $\frac{\delta^2}{N-1}(wb^*+c)$ at $\ell = 0$ and $\frac{\delta^2}{N-1}(w+c)-(s-w)$ at $\ell = 1$. At $\ell = 0$ the LHS is zero and the constraint is automatic; the binding is therefore at $\ell = 1$ and reduces to

$$s-w \;\le\; \frac{\delta^2}{N-1}(w+c), \tag{A.7}$$

the second inequality in (9). This yields $N^*(1) = 1+\delta^2(w+c)/(s-w)$.

Combining Steps 1–6, a maximal-trust hierarchical equilibrium at $J = 1$ under the family (8) exists if and only if $s < w/(1-\delta)$ and $s-w \le \delta^2(w+c)/(N-1)$, and is characterised by (v^*, b^*) of (10).

A.4 Proof of Proposition 3

The proof follows the same logic as Proposition 1, but requires three adjustments reflecting selective trade: (i) payoffs scale with J for active agents, (ii) deviations are evaluated relationship-by-relationship, and (iii) continuation values depend on the full seniority profile.

Step 1 (Necessity). Fix a profile $\boldsymbol{\ell}$ and an active agent i with $\ell_i \geq \hat{\ell}(\boldsymbol{\ell})$. On the candidate path she serves J principals and earns $J\ell_i w + \delta v(\ell_i, \boldsymbol{\ell}_{-i})$.

(a) Cheat-and-leave. If she defects against all J principals and migrates, she obtains $J\ell_i s + \delta v(0)$. The deviation is not profitable iff

$$J\ell_i(s - w) \leq \delta\big(v(\ell_i, \boldsymbol{\ell}_{-i}) - v(0)\big),$$

which is (IC_L^{TS}). Maximal trust implies equality for all active interior types.

(b) Cheat-and-stay. Deviations are bilateral: defecting against one principal does not affect the others. For any given relationship, the one-period gain is $\ell_i(s-w)$. The punishment is one-period Nash reversion, applied only if the pair survives and is rematched. For a top-ranked agent, the rematch probability is $J/(N-1)$, so the deviation is not profitable iff

$$\ell_i(s - w) \leq \frac{\delta^2 J}{N-1}(w + c).$$

Since the left-hand side is increasing in ℓ_i while the right-hand side is constant, the constraint binds at $\ell_i = 1$, yielding (IC_R^{TS}).

Step 2 (Sufficiency: on-path deviations). Under (IC_L^{TS})–(IC_R^{TS}), no active agent has a profitable one-stage deviation. Inactive agents have no current match and therefore no profitable deviation. As in Proposition 1, principals optimally choose $\lambda = \ell$.

Step 3 (Off-path continuation). Consider a one-stage deviation in which agent i defects against one principal. The candidate continuation prescribes that the principal returns while the agent leaves.

Let $V^R_{\text{after}}(\ell_i, \boldsymbol{\ell}_{-i})$ denote the continuation value from returning after cheating. The cheat-and-return constraint implies

$$\ell_i(s - w) \leq \delta\big(v(\ell_i, \boldsymbol{\ell}_{-i}) - V^R_{\text{after}}(\ell_i, \boldsymbol{\ell}_{-i})\big),$$

while maximal trust implies that the cheat-and-leave constraint binds:

$$J\ell_i(s-w) = \delta\big(v(\ell_i, \boldsymbol{\ell}_{-i}) - v(0)\big).$$

Combining the two inequalities yields

$$V^R_{\text{after}}(\ell_i, \boldsymbol{\ell}_{-i}) \leq v(0).$$

Thus, once the current gain is sunk, the deviating agent weakly prefers leaving to returning.

The principal's return is optimal given that the agent leaves and that future punishment opportunities are preserved.

Step 4. The prescribed one-period punishment is sequentially rational within the bilateral relationship. By the one-stage deviation principle, these checks establish that the candidate profile is a sequential equilibrium.

A.5 Proof of Proposition 4

Fix $J \geq 2$. Let $\hat{\ell}(\boldsymbol{\ell})$ denote the finite-N activity cutoff and let $\bar{\ell}$ satisfy $1 - F(\bar{\ell}) = \frac{1}{J}$.

Step 1. Convergence of the cutoff. Let $\hat{\ell}(\boldsymbol{\ell})$ denote the empirical cutoff selecting the top fraction $1/J$ of the population, i.e.

$$1 - F_N(\hat{\ell}(\boldsymbol{\ell}); \boldsymbol{\ell}) = \frac{1}{J}.$$

Fix $\varepsilon > 0$. Since F is continuous at $\bar{\ell}$ and satisfies

$$1 - F(\bar{\ell}) = \frac{1}{J},$$

we have

$$F(\bar{\ell} - \varepsilon) < 1 - \frac{1}{J} < F(\bar{\ell} + \varepsilon).$$

By uniform convergence of F_N to F, with probability approaching one,

$$F_N(\bar{\ell} - \varepsilon; \boldsymbol{\ell}) < 1 - \frac{1}{J} < F_N(\bar{\ell} + \varepsilon; \boldsymbol{\ell}).$$

By monotonicity of F_N, this implies

$$\bar{\ell} - \varepsilon < \hat{\ell}(\boldsymbol{\ell}) < \bar{\ell} + \varepsilon.$$

Hence

$$\hat{\ell}(\boldsymbol{\ell}) \xrightarrow{p} \bar{\ell}.$$

Step 2. Convergence of the flow payoff. Fix ℓ_i. Define the finite-N flow payoff

$$r_N(\ell_i, \boldsymbol{\ell}_{-i}) = \int_{[0,1]^N} J\tilde{\ell}_i w\, \mathbf{1}\{\tilde{\ell}_i \geq \hat{\ell}(\tilde{\boldsymbol{\ell}})\}\, \mathcal{G}(\tilde{\boldsymbol{\ell}}, \boldsymbol{\ell})\, d\tilde{\boldsymbol{\ell}},$$

and the limiting flow payoff

$$r(\ell_i) = \int_0^1 J\tilde{\ell}_i w\, \mathbf{1}\{\tilde{\ell}_i \geq \bar{\ell}\}\, g(\tilde{\ell}_i, \ell_i)\, d\tilde{\ell}_i.$$

The difference between r_N and r is driven only by the difference between $\hat{\ell}(\tilde{\boldsymbol{\ell}})$ and $\bar{\ell}$. Since $\hat{\ell}(\tilde{\boldsymbol{\ell}}) \to \bar{\ell}$ in probability and the event $\{\tilde{\ell}_i = \bar{\ell}\}$ has probability zero under $g(\cdot, \ell_i)$, it follows that

$$r_N(\ell_i, \boldsymbol{\ell}_{-i}) \xrightarrow{p} r(\ell_i).$$

Step 3. Convergence of the value function. The finite-N value function satisfies

$$v_N(\ell_i, \boldsymbol{\ell}_{-i}) = r_N(\ell_i, \boldsymbol{\ell}_{-i}) + \delta \int_{[0,1]^N} v_N(\tilde{\ell}_i, \tilde{\boldsymbol{\ell}}_{-i})\, \mathcal{G}(\tilde{\boldsymbol{\ell}}, \boldsymbol{\ell})\, d\tilde{\boldsymbol{\ell}},$$

while the limiting value function satisfies

$$v(\ell_i) = r(\ell_i) + \delta \int_0^1 v(\tilde{\ell}_i)\, g(\tilde{\ell}_i, \ell_i)\, d\tilde{\ell}_i.$$

Both operators are contractions with modulus δ, and payoffs are uniformly bounded. Standard arguments for perturbed contractions imply that convergence of the flow payoff $r_N \to r$ implies convergence of the fixed points. Hence

$$v_N(\ell_i, \boldsymbol{\ell}_{-i}) \xrightarrow{p} v(\ell_i)$$

for every fixed ℓ_i.

Step 4. Continuity of the limit. The limiting Bellman operator maps continuous functions into continuous functions: the kernel g is continuous, and the cutoff affects the integrand only on a set of measure zero. Since the operator is a contraction, its unique fixed point v is continuous.

A.6 Proof of Proposition 5

The argument parallels Proposition 1. The only difference is that the cost of restarting is the entry investment m rather than the loss of seniority. A cheat-and-leave deviation yields gain $s-w$ today and requires paying m upon re-entry. It is therefore unprofitable if and only if

$$s - w \le \delta m,$$

which is (IC_L^I).

A cheat-and-return deviation is bilateral and identical to the $J = 1$ case. The deviation is unprofitable if and only if

$$s - w \le \frac{\delta^2}{N-1}(w + c),$$

which is (IC_R^I).

Combining the two constraints implies that, after cheating, the agent weakly prefers leaving to returning, so the prescribed off-path continuation is sequentially rational. The punishment is the stage-game Nash outcome.

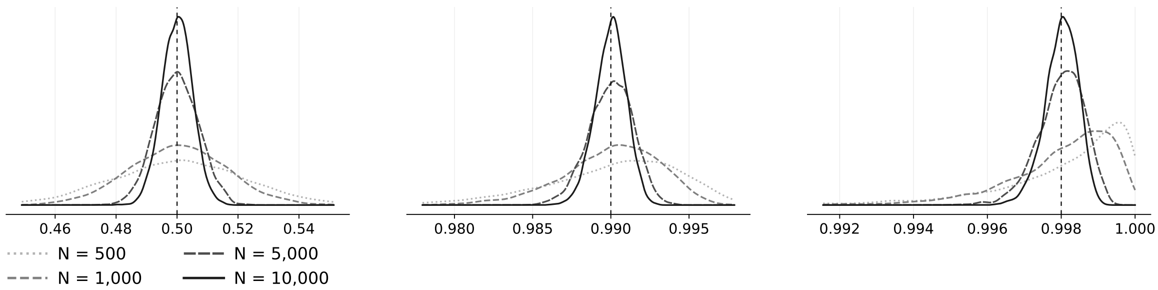
Setting (IC_L^I) at equality gives the minimal investment

$$\underline{m} = \frac{s-w}{\delta},$$

and the entrant's net payoff follows.

A.7 Numerical validation of the cutoff approximation

We quantify the convergence of the empirical activity cutoff $\hat{\ell}(\boldsymbol{\ell})$ to its deterministic limit $\bar{\ell}$. On the candidate path the cross-section is i.i.d. across individuals (deaths and advancements are independent across i), so we draw N stationary seniorities directly and verify against a T=2,000 time-series chain. For each $J \in \{2, 100, 500\}$ and $N \in \{500,\ 1{,}000,\ 5{,}000,\ 10{,}000\}$ we form R=5,000 communities and compute $\hat{\ell}(\boldsymbol{\ell})$.



(a) $J = 2$, $\bar{\ell} = 0.500$. (b) $J = 100$, $\bar{\ell} = 0.990$. (c) $J = 500$, $\bar{\ell} = 0.998$.

Figure 5: **Empirical activity cutoff concentrates on the deterministic limit.** Density of $\hat{\ell}(\boldsymbol{\ell})$ across R = 5,000 communities; vertical dashed line marks $\bar{\ell}$. *Primitive:* the kernel $g(\tilde{\ell}, \ell)$ of equation (7) with $(s, w, \delta) = (2, 1, 2/3)$ and the equilibrium-pinned mobility $b = b^* = 1 - (s - w)/(\delta s) = 0.250$. The induced steady-state CDF is $1 - F(\ell) = (1 - \ell)^\alpha$ with $\alpha = (1 - b^*)(1 - \delta)/b^* = 1$, so $\bar{\ell} = 1 - 1/J$. Legend in panel (a) applies throughout: lines from light dotted to dark solid correspond to $N \in \{500,\ 1{,}000,\ 5{,}000,\ 10{,}000\}$.

At this calibration the empirical cutoff is unbiased and tightly concentrated. For $J = 2$, $\text{sd}(\hat{\ell})$ falls from 0.022 at N=500 to 0.016 at N=1,000, 0.0071 at N=5,000, and 0.0050 at N=10,000; for $J = 100$ from 0.0044 to 0.0010 over the same range; for $J = 500$ from 0.0019 to 0.00045. The mean of $\hat{\ell}$ is indistinguishable from $\bar{\ell}$ at every (J, N), and the contraction follows the standard $1/\sqrt{N}$ rate expected for order statistics.

B The three-individual finite case

This appendix derives the smallest non-trivial finite case, $N = 3$ with $J \in \{1, 2\}$, and verifies both incentive constraints in closed form at a stationary candidate profile. The derivation shows that the within-group return constraint under $J = 2$ admits an explicit range of primitives (s, w, δ, c) for which a hierarchical equilibrium exists that is not supported under $J = 1$ at the same primitives. It thus corroborates the $J > 1$ contrast of Section 3.3 at the smallest finite community size at which the capacity constraint can bind.

Primitives and candidate profile. The community hosts three members; $N = 3$, $M \to \infty$ as in the main text. The seniority vector $\boldsymbol{\ell}_t = (\ell_{[1]t}, \ell_{[2]t}, \ell_{[3]t})$ is the ordered triple from lowest to highest. Every surviving member advances along the kernel (8) and newborns enter at a draw from $G(\cdot, 0)$. We specialise to the discretised rank set $L_3 = \{\ell_a, \ell_b, \ell_c\}$ with $0 < \ell_a < \ell_b < \ell_c = 1$, with advancement one step up per period, conditional on survival, as the finite analogue of the one-parameter kernel: this is the simplest transition structure that (i) satisfies Assumption 1, (ii) has an absorbing top rank, and (iii) admits a stationary distribution. The candidate profile is maximal-trust hierarchical: principals extend $\lambda = \ell_q$ to an agent at ℓ_q, matched agents play `cooperate`, and every surviving member stays.

Case $J = 1$ (every member active). At $J = 1$ each of the three members is both a principal and an agent in exactly one pair per period. By Proposition 1 the value function is $v(\ell) = w/(1-\delta) - (1-\ell)(s-w)/\delta$, and the binding return constraint at $\ell = 1$ reads

$$\mathrm{IC}_R^{N=3,J=1}: \qquad s - w \ \le \ \frac{\delta^2(w+c)}{N-1}\Big|_{N=3} \ = \ \frac{\delta^2(w+c)}{2}. \tag{B.1}$$

The migration constraint IC_L^H at $\ell = 1$ reduces to $s - w \le \delta(v(1) - v(0)) = w \cdot \delta \cdot [1 - (1+\alpha \cdot 0)/(1+\alpha)] = w - w/(1+\alpha)$, which under $b^* = 1 - (s-w)/(\delta s)$ is satisfied with equality in the maximal-trust selection and with strict slack whenever $\alpha < \infty$, i.e., whenever $\varepsilon < \varepsilon_0$.

Case $J = 2$ (only top rank active). At $J = 2$ the top rank $\ell_c = 1$ is active; she serves both other members as principals in the same period. Only one agent position is filled; the two lower ranks are inactive as agents. The top rank earns $Jw = 2w$ per period and, by (P3), $v(1) = 2w/(1-\delta)$. Her return constraint from a defect-and-stay deviation, by (P4) and Step 7 of A.2 specialised to $N = 3$, reads

$$\mathrm{IC}_R^{N=3,J=2}: \qquad s - w \ \le \ \frac{\delta^2 J(w+c)}{N-1}\Big|_{N=3,J=2} \ = \ \delta^2(w+c), \tag{B.2}$$

a factor-of-two relaxation relative to (B.1). The migration constraint at $\ell = 1$ compares the staying stream $v(1) = 2w/(1-\delta)$ against the defect-and-migrate stream $2s + \delta v^{\text{newcomer}}$, where v^{newcomer} is the continuation value of a newcomer entering the receiving community at a draw from $G(\cdot, 0)$. Under the discretised L_3 kernel with absorbing top and one-step upward advancement, $v^{\text{newcomer}} \le v(1)$ with strict inequality, and the constraint $2(s-w) \le \delta[v(1) - v^{\text{newcomer}}]$ is equivalent to the active-rank migration constraint IC_L^{TS} of Equation 17 at $\ell = 1$; direct computation gives strict slack whenever (B.2) holds, because the staying continuation $v(1)$ scales with $J = 2$ while the defect-and-migrate continuation v^{newcomer} does not.

Existence contrast and admissible range. Combining the two cases, define the primitives $(\varepsilon, \delta, c/w)$ with $\varepsilon = (s - w)/w \geq 0$ and $c/w \geq 0$. The $J = 1$ region is $\varepsilon \leq (\delta^2/2)(1 + c/w)$; the $J = 2$ region is $\varepsilon \leq \delta^2(1 + c/w)$. The *HE-gap region* is the set of primitives admissible at $J = 2$ but not at $J = 1$:

$$\frac{\delta^2}{2}(1 + c/w) \; < \; \varepsilon \; \leq \; \delta^2(1 + c/w). \tag{B.3}$$

For $\delta = 0.9$, $c/w = 0$ the gap is $\varepsilon \in (0.405, 0.810]$; for $\delta = 0.9$, $c/w = 1$ the gap is $\varepsilon \in (0.810, 1.620]$. Both are non-empty intervals of admissible temptations, and both intersect the empirically relevant range $\varepsilon \in (0, 1)$. In this HE-gap region, a hierarchical equilibrium with selective trade exists at $J = 2$ and $N = 3$ but not at $J = 1$ and $N = 3$—the $N = 3$ analogue of part (ii) of Corollary 2.

Comparative statics. Three comparative statics of the $N = 3$ case carry over to the general mean-field analysis: (i) the gap region widens as c/w rises (because the RHS of both constraints scales with $1 + c/w$ and the gap scales linearly), consistent with $\partial N^*/\partial c > 0$ in Table 1; (ii) the gap region widens as δ rises (quadratically), consistent with $\partial N^*/\partial \delta > 0$; (iii) the gap region contracts as ε rises with fixed δ, consistent with $\partial N^*/\partial(s - w) < 0$. The $N = 3$ case therefore reproduces the direction and sign of the mean-field comparative statics at the smallest finite community size at which the capacity constraint can bind, corroborating that the core-regime conclusions of Section 3.3 are not an artefact of the large-population limit.

C Qualitative applications

The Roman Republic of Section 4 is the case in which every primitive of the model has a codified institutional counterpart. Three further settings sharpen the same comparison from Sections 3.5–3.4: a hierarchical equilibrium with selective trade (HE-TS), a reciprocity-based arrangement with continuation $w/(1-\delta)$, and an identity-investment equilibrium (IIE) sustained by a sunk entry cost m. Throughout, J denotes capacity (the number of trading slots a senior agent fills per period), $\boldsymbol{\ell}_t$ the publicly observable rank vector at date t, and ε the size of the migratory outside option relative to the cooperative continuation; (DE) is the dynamic-evolutionary stability condition of

Proposition 6, that states that the active-rank value, and the destabilisation of the reciprocity-based comparator under a rise in ε. Also Proposition 6, delivers the parallel destabilisation of the IIE comparator. The cases are pre-state hierarchy among non-agrarian hunter–gatherers, the coexistence of Muslim and Maguzawa Hausa in twentieth-century Nigeria, and seniority-based cooperation in online communities; each isolates a distinct empirical wedge — C.1 the heterogeneity-free margin, C.2 the entry-direction wedge, C.3 the maintainership-transfer asymmetry.

C.1 The birth of hierarchy in early human societies

Prior to the introduction of settled agriculture around 12,000 years ago, human societies consisted of small groups and population densities as a whole were relatively low [Boix and Rosenbluth, 2014, Keeley, 1988]. They also tended to be egalitarian (Boix and Rosenbluth [2014]). Inequality increased, and average health appeared to worsen, however, after the introduction of agriculture [Boix and Rosenbluth, 2014, Bowles and Choi, 2013], a period that also saw an increase in population densities and the emergence of social hierarchies [eg Boehm, 1999, Price and Brown, 1985].[26] Some scholars argue that egalitarianism was pinned down by the "countervailing power" of weaker individuals– particularly their ease of exit, should any 'self-aggrandizing' individual seek to dominate [Woodburn, 1982, Knauft, 1991, Boehm, 1999, Nowak and Sigmund, 2005, Sterelny, 2013, Seabright, 2013, Bowles and Choi, 2013].

An important body of work links the emergence of hierarchies and within-group inequality to the development of settled agriculture.[27] A series of important papers show how hierarchies can emerge as an endogenous response that favors some individuals or groups over others, often through mechanisms that relate to heterogeneity among individuals.[28]

[26] Summarizing a large body of anthropological and archaeological work, the anthropologist Robert Boehm [1999][pg3-4] states: "before twelve thousand years ago, humans basically were egalitarian. They lived in what might be called societies of equals, with minimal political centralization and no social classes. Everyone participated in group decisions and outside the family there were no dominators."

[27] See also Michalopoulos and Papaioannou [2020] who show that pre-colonial jurisdictional hierarchy and political centralization are strongly associated with contemporary regional development in Africa.

[28] This literature establishes that ecological conditions, appropriability, infrastructure, and trade routes can generate hierarchical political organization and that such hierarchy has persistent economic consequences. For example, Flückiger et al. [2024] point to the importance of metal trade, transit bottlenecks, and the emergence of taxing elites, as a driver of hierarchy. Mayshar et al. [2022] show that hierarchy and early states arose where staple crops were more appropriable—cereals rather than roots and tubers—so that elites could extract surplus, while Dow and Reed [2013] model the emergence of hereditary elites and commoners from productivity growth, enclosure of high-quality sites, and insider control over land. Allen et al. [2023] instead distinguish cooperative from extractive theories of

Our contribution is complementary but distinct: rather than identifying a particular historical shock that creates elites, we provide a mechanism for why hierarchical communities can sustain trust once outside options and mobility make "cheat-and-leave" deviations attractive. In our model, public seniority ranks coexist with private bilateral histories; when capacity exceeds one, selective trade concentrates high-stake interactions on senior members, makes cooperation scalable, and gives hierarchical communities dynamic robustness against both reciprocity-based and identity-investment alternatives—properties that the unit-capacity hierarchy does not deliver. Further, our model suggests that such heterogeneity is not necessary for inequality to develop. Any exogenous change that makes it possible to do more tasks for others, *even if this affects all individuals equally*, could lead to the development of hierarchies. For example, hierarchies might emerge from increases in population density, that make alternative trading partners and communities more accessible, or technological improvements that reduce travel costs.

Indeed, Figures 6 and 7 reveal a positive relationship between hierarchy and population density in a "comprehensive" dataset of modern (non-agricultural) hunter-gatherer societies, as observed upon their first contact with Europeans [Binford, 2001][pg.117]. Consistent with our theory, egalitarian hunter-gatherer ethnic groups tend to be much smaller, even if we include within them those with some form of institutionalized leadership (Figure 6). Further, (Figure 7) while egalitarian institutions survive with low population densities, even among hunter-gatherer societies, ethnic groups with wealth distinctions and social stratification begin to be more prevalent when population densities exceed as little as 1 person per square kilometer.

Our model also sheds new light on a lingering paradox: that hierarchies and inequalities *precede* the development of early states with coercive power that might justify having large groups for mutual defence, or indeed might result in the security of property rights that would allow 'self-aggrandizing' individuals to assert economic distinction in the first place [Seabright, 2013].[29] The hierarchical equilibria that emerge in our model address

government, showing in southern Iraq that river shifts created demand for public canal coordination while reducing the scope for purely local extraction; Bentzen et al. [2017] link irrigation dependence to autocracy and land inequality through elite control over water and arable land Dow and Reed [2013] present a theoretical model where geographical heterogeneity in agricultural suitability generates inequality, as early movers to desirable plots of land collude to exclude latecomers from property ownership. See also Boix [2015], who argues that agricultural technology favoured some agents more than others, leading the latter to become bandits and then specialists in violence.

[29] In his survey of archaic states, Norman Yoffee [2005][pg.35], based upon archaeological evidence such as differing sizes of residences, the distribution of artifacts, and mortuary furnishings, concludes that "no prehistoric trajectory to any state fails to contain indications of significant economic inequality or the potential of such inequality well before the appearance of anything that might be called a state."

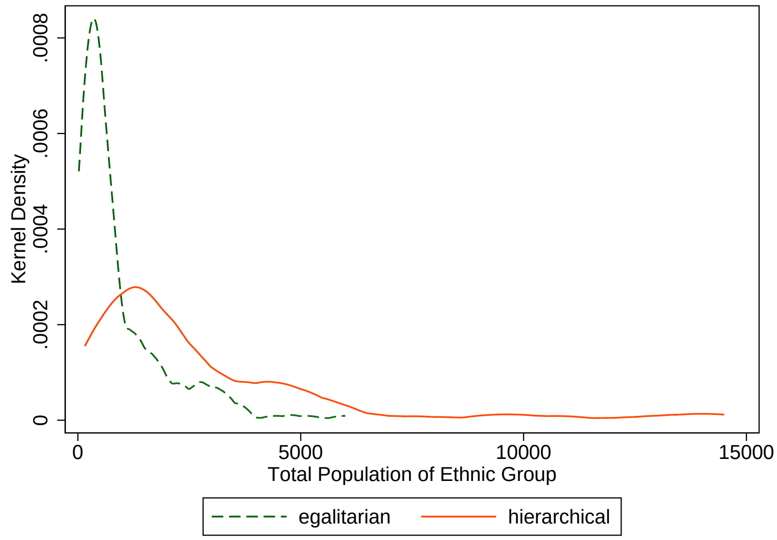


Figure 6: **Egalitarian Hunter-Gatherer Ethnic Groups are Smaller**

Kernel densities of ethnic group size are based upon the Binford [2001] dataset of 339 hunter-gatherer societies described at first contact with anthropologists. This viewed as a comprehensive and expanded version of those mentioned in Murdock's *Ethnographic Atlas*. *Egalitarians* include all societies classified as "generic hunter-gatherers" or "generic hunter-gatherers with institutionalized leaders". *Hierarchical* are all societies classified as "wealth-differentiated hunter-gatherers" or "stratified or characterized by elite and privileged leaders".

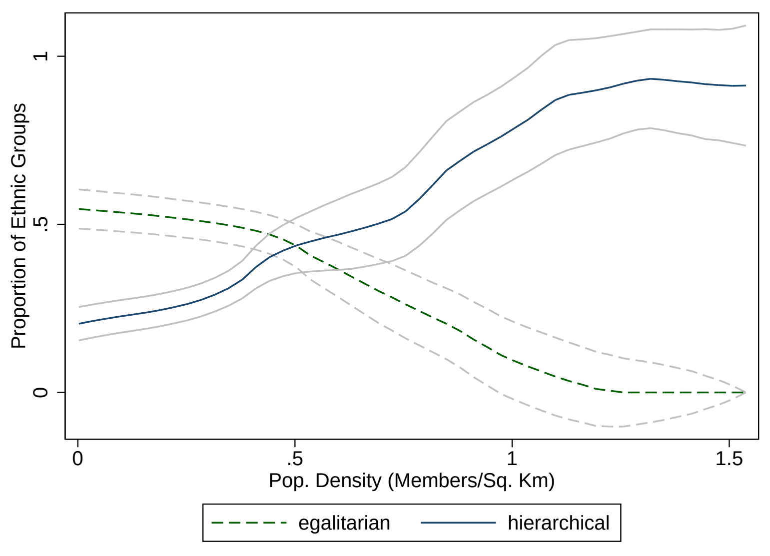


Figure 7: **Egalitarian Groups Do Not Survive as Population Densities Rise**

Local moving averages (epanechnikov kernel, ROT bandwidth) showing the proportion of societies that are egalitarian or hierarchical at different levels of ethnic group population densities. These are based upon the Binford [2001] dataset of 339 hunter-gatherer societies described at first contact with anthropologists. This viewed as a comprehensive and expanded version of those mentioned in Murdock's *Ethnographic Atlas*. *Egalitarians* include all societies classified as "generic hunter-gatherers" or "generic hunter-gatherers with institutionalized leaders". *Hierarchical* are all societies classified as "wealth-differentiated hunter-gatherers" or "stratified or characterized by elite and privileged leaders".

the question of why elites emerge: while egalitarian societies are favored when distances or travel costs are great and when the tasks individuals can perform for one another are also extremely limited, hierarchies are favored when these capacity constraints are relaxed. Furthermore, in environments where trade can be concentrated with senior members, hierarchies can be designed that are better able to sustain cooperation in large groups, and are robust to *countervailing power*- coalitional deviations. The development of social distinction also appears natural and intuitive- the hierarchies that emerge transfer surplus from newcomers to incumbent (returning) members of society, yet sustain incentives for newcomers to themselves return. Our model also explains the key question of institutional selection: hierarchies can emerge spontaneously that dominate and lead to the breakdown in other institutional arrangements, and even undermine groups that had previously been able to maintain such cooperation without institutions. Thus, even without any additional assumptions that hierarchies enjoy a technological advantage in coercive organization, hierarchies may emerge spontaneously that sustain cooperation, and may have provided the seed of early political states.[30]

C.2 Pastoralist and agriculturalist Hausa in Nigeria

The case of the Hausa in Nigeria further illustrates how emergent hierarchies sustain trust in a modern society with weak legal institutions while undermining cooperation

[30] More formally: a small foraging band corresponds to a community at small N in which cooperation is sustained on a random-matching candidate path with continuation value $w/(1-\delta)$ — the reciprocity-based comparator of Section 3.5. The band sits at the unit-capacity benchmark $J = 1$ of Section 3.2, with ceiling $N^*(1) = 1+\delta^2(w+c)/(s-w)$, independent of the matching kernel. The key comparison is with institution-free communities: egalitarianism is documented as kin- and reciprocity-based rather than gated by an entry payment. In the early hierarchical settings, by contrast, the public seniority record $\boldsymbol{\ell}_t$ is implemented through ritual office, residence size, mortuary furnishing, and visible insignia of rank [Yoffee, 2005]; bilateral matchings remain anonymous in the sense of Section 2, so that the public seniority record coexists with anonymous pairwise interaction. The settings of interest sit at $J \geq 2$, with selective trade implemented through the concentration of trade, redistribution, and ritual office in a small number of high-ranked individuals.

As population density rises, mobility across bands becomes cheaper and the migratory outside option $s + \delta\, w/(1-\delta)$ grows relative to the cooperative continuation; ε enters the working range of (DE) on which Proposition 6 holds and the reciprocity-based path unravels, while members at active hierarchical ranks earn a continuation strictly above $w/(1-\delta)$ (Part B, the active-rank value) and prefer to remain. The discriminating empirical prediction, which notably does not require agent heterogeneity, is that hierarchy emerges in cross-sectionally homogeneous populations once density crosses the threshold — precisely the Binford [2001] pattern, in which hierarchy share rises with density and not with any plotted measure of within-group heterogeneity. Thus, the HE-TS supplies an institutional selection mechanism that complements existing accounts, without *requiring either* heterogeneity or from a specific coincident technological prime mover. Instead the possibility of increased capacity $J \geq 2$ is the key operative margin and the cross-section pattern in Binford [2001] as the empirical shadow of Part C.

in nearby egalitarian communities.[31]

Among the Hausa, it is religious identity– particularly Islam– that plays an important role in shaping access to commercial opportunities. However, Hausa Muslims and non-Muslims (known as Maguzawa) have common ethnic origins and have lived intermixed historically, with Maguzawa seen by Muslims as being the "original" Hausa [Last, 1999].[32] Though Maguzawa now constitute fewer than 2% of the 50 million modern Hausa-speakers, mass conversion to Islam is relatively recent: in 1900, only 5% of contemporary Hausaland were Muslim [Salamone, 2010].

Yet, despite their similarities, and the ability of individuals to move relatively freely between Muslim and Maguzawa groups within the Hausa, there remain remarkable differences in their social organization and levels of trust. The Muslim Hausa have a well-developed hierarchical structure that transcends lineage, while the Maguzawa have well-established norms of egalitarianism [King, 2006, Last, 1979, 1999, Barkow, 1974, Salamone, 2010]. As in the hierarchical equilibrium, access to enhanced trading opportunities is directly related to the extent to which the agent is embedded in a network of trust-based relations. Those with the highest status include long-distance Muslim traders, who traded along pilgrimage routes across the Sahara [King, 2006, Barkow, 1974, Cohen, 1969]. Those involved in smaller-scale commerce or professions involving less trust also have lower perceived status [Barkow, 1974]. Qualitative accounts stress how among the Hausa, "a good deal of business is conducted with handshakes and one's word" [Salamone, 2010][p.3].

In contrast to the Muslim Hausa, the Maguzawa are a good example of an 'institution-free' competing group. Maguzawa have strong norms of egalitarianism and reciprocal obligation [King, 2006, Last, 1979, 1999, Barkow, 1973, Salamone, 2010]. They are also perceived as being completely open and welcoming to new entrants [Barkow, 1973, Last, 1999]. Though, the Maguzawa have lived in close proximity to the Muslims for centuries, there is no ambiguity in classification: both groups maintain clear (though reversible) visual symbols of distinction, including different styles of clothing and restrictions on women [Barkow, 1973, Last, 1999].

[31] Of the 20 sub-Saharan African states surveyed by Afrobarometer in 2008, Nigerians ranked last, with only 9.50% stating that they trusted non-relatives they knew "a lot", compared to 25.50% of other sub-Saharan African respondents. Formal third party enforcement is also perceived to be among the weakest in sub-Saharan Africa.

[32] As Last [1979][239] notes, the Maguzawa were "considered one of "us", therefore not targets for jihad even in the early nineteenth century . . . There was apparently, and still is, no urge among the Muslims to convert the non-Muslims among their neighbors and subjects . . . "

Table 2: **Trust in Nigeria by ethnicity and religion**
Proportion of population that trust other Nigerians "somewhat" or "a lot" (as opposed to "not at all" or "just a little".) source: Afrobarometer 2008-2009

		Non-Muslims	Muslims	All
Proportion claiming to trust:				
Non-Hausa	Relatives	0.678	0.741	0.692
	Others they know	0.386	0.464	0.403
	Other Nigerians	0.247	0.289	0.257
	Observations	1,394	401	1,795
Hausa	Relatives	0.918	0.852	0.858
	Others they know	0.469	0.635	0.620
	Other Nigerians	0.367	0.469	0.459
	Observations	49	480	529

Hausa thus have had the option of choosing to be Muslim and thus joining the hierarchical structure of the Muslim Hausa, or choosing to remain or even switch to being Maguzawa [Last, 1979].[33] Consistent with the model, new converts to Islam are *actually poorer* than non-Muslim Hausa [Last, 1979, 1999, Barkow, 1974].[34] Yet, despite the drop in wealth, it appears that it is the possibility of advancement in the Muslim hierarchy (and thus of being the focus of coordinated trades and thus increased wealth in the future), that drives individuals to convert.[35]

Further, despite common ethnic, linguistic and geographic endowments, and the relative lack of barriers to conversion to Islam [Last, 1979], the differences in the degree of trust between Hausa religions are remarkable. Table 2 shows the proportion of Nigerian respondents to the Afrobarometer survey who reported their willingness to trust others from their country "a lot" or "somewhat". As the table suggests, while both non-Muslim and Muslim Hausa tend to report a greater willingness to trust others in general compared to non-Hausa Nigerians, there are remarkable differences within the Hausa according to the degree they trust people within and outside their own families. While a greater proportion of non-Muslim Hausa show a willingness to trust relatives, Muslim Hausa are more likely to report that they trust both non-relatives that they know and strangers more generally.

[33] In fact, consistent with the model, new Muslim converts who "fail" in trade can and do change their identity to Maguzawa [Last, 1999].

[34] Ironically, in the early years of independence, due to their relatively high wealth, the agriculturalist Maguzawa were classed as "merchants" and taxed at a higher rate on average [Last, 1999].

[35] Last [1979] describes: "Though it is clear to non-Muslims that the Muslim is often poorer and lives a more constricted social life, the prospect is of wealth through trade . . . (239)"

Consistent with our model, rather than the 'institution-free' community undermining cooperation among the Muslim Hausa, the Muslim community has attracted more converts, despite the fact that converts tend to be poorer on average, and it has been cooperation among the Maguzawa that has largely broken down. Once living in close proximity to Muslims in the towns of Hausaland, the remaining Maguzawa have become specialized economically in largely autarkic farming activities, and have become more geographically dispersed, both from Muslims and from each other [Greenberg, 1947, Last, 1979, 1999].[36]

In the language of our model, the Muslim Hausa correspond to HE-TS: status is publicly observable through religious identity and trading reputation, the highest ranks are occupied by long-distance traders for whom $J \geq 2$ is the empirically relevant case, trade is selectively concentrated toward those ranks, and "a good deal of business is conducted with handshakes and one's word" [Salamone, 2010, p. 3]. The seniority record $\boldsymbol{\ell}_t$ is implemented through religious identity, position in the trans-Saharan trading hierarchy, and public observance, while bilateral matchings remain anonymous in the sense of Section 2, as in C.1. The Maguzawa correspond to a reciprocity-based arrangement in the sense of Section 3.5: cooperation is sustained on a random-matching path among kin and neighbours, and an active member's continuation value is $w/(1-\delta)$. The newcomer-versus-active-rank wedge resolves the puzzle of why Maguzawa would prefer to join the Muslim community despite the fact that newcomers are poorer. The newcomer value $v^*_{\mathrm{MF}}(0;\varepsilon)$ in HE-TS lies below $w/(1-\delta)$ at every $\varepsilon > 0$ (Part A, the newcomer value), yet a member at an active rank earns strictly more than $w/(1-\delta)$ (Part B, the active-rank value); the same agent therefore prefers HE-TS in expectation through the channel of advancement up the trading hierarchy [Last, 1979, p. 239], not through a contemporaneous wealth comparison. Part C predicts the destabilisation of the reciprocity comparator once ε enters the working range of (DE), and Part C′ of Proposition 6 delivers the parallel destabilisation of the identity-investment comparator; the joint statement has an empirical shadow: in towns of Hausaland above two hundred per square mile, the Maguzawa population dwindles [Last, 1979, 1999, Greenberg, 1947]. The mechanism for one-sided conversion is then the dynamic value of advancement: converts arrive poorer and stay because the expected continuation, not contemporaneous payoff, governs entry.

[36] In the 1970s, virtually all Hausa in areas with population densities above 200 per square mile had become Muslim [Last, 1979].

C.3 Online communities

Online communities — open-source software (OSS) projects, technical discussion fora, peer-review platforms — sustain cooperation at scale without monetary compensation, in the presence of substantial anonymity, and often after displacing earlier reciprocity-based arrangements (mailing lists, single-firm proprietary alternatives) that had been the dominant venues for the same activity [Raymond, 2001, Lerner and Tirole, 2005, Shah, 2006]. The key puzzle is again how cooperation can be sustained under conditions that make both bilateral monitoring and symmetric reputational punishment difficult.[37]

Hierarchies are once again common in this setting. Seniority is publicly observable and codified: OSS projects often designate a restricted set of senior contributors, "committers," with the right to merge code into the main branch; many fora display the date of registration and a count of past contributions next to each post. The seniority record $\boldsymbol{\ell}_t$ has a literal implementation as a public profile attribute, and the underlying matchings between contributors and reviewers remain anonymous in the sense of Section 2, the most literal instance of the public-rank/anonymous-match distinction in the three cases. Capacity is genuinely binding: a senior contributor's time available to review patches, mentor newcomers, or maintain a release branch is bounded, so $J \geq 2$ is empirically relevant and trade is selectively concentrated toward high-seniority members. The per-match payoffs have direct counterparts: w is the surplus from a useful contribution successfully integrated and supported, $s - w$ the short-run gain from low-effort contributions or from withholding support, c the exposure cost of integrating an unvetted patch.

A key discriminating prediction of our model is on the direction of *maintainership transfer*: when a senior committer exits, HE-TS predicts succession by an internal contributor of high $\boldsymbol{\ell}_t$ rather than by an external entrant of equivalent contemporaneous skill, because the active-rank value (Part B) lies strictly above the migratory outside option for the senior cohort and below it for an outside replacement. These patterns indeed appear to mimic the qualitative record that has emerged for mature open source software communities [Raymond, 2001, Lerner and Tirole, 2005, Shah, 2006, Eghbal, 2020].[38]

[37]Beyond the literature already surveyed, we complement the contemporary governance literature on OSS [Eghbal, 2020, Dey et al., 2020] documents the role of committer status and maintainership in concentrating effort, and Tirole [2021] models scoring and ranking on digital platforms through rating aggregation rather than capacity-bounded selective trade.

[38]A number of important alternative theoretical accounts predict otherwise: network-density mechanisms point to the most-connected contributor, reputation-system mechanisms to the highest-rated outsider, and intrinsic-motivation accounts are silent on the asymmetry. .

References

George A. Akerlof and Rachel E. Kranton. Economics and identity. *Quarterly Journal of Economics*, 115(3):715–53, 2000.

S. Nageeb Ali and David A. Miller. Ostracism and forgiveness. *American Economic Review*, 106(8):2329–2348, August 2016.

Robert C. Allen, Mattia C. Bertazzini, and Leander Heldring. The economic origins of government. *American Economic Review*, 113(10):2507–2545, October 2023.

Aristotle. *Politics*. Chicago U. Press, Chicago, translated by Carnes Lord, 2013 edition, 335BC.

Kenneth J. Arrow. *The limits of organization*. Norton, New York, NY, 1st edition, 1974.

M.M. Austin, editor. *The Hellenistic World from Alexander to the Roman Conquest. A selection of ancient sources in translation*. Cambridge University Press, 2nd edition, 2006.

Jerome H. Barkow. Muslims and maguzawa in north central state, nigeria: an ethnographic comparison. *Canadian Journal of African Studies*, 7(1):59–76, 1973.

Jerome H. Barkow. Evaluation of character and social control among the hausa. *Ethos*, 2(1):1–14, Spring 1974.

Robert Bates, Avner Greif, and Smita Singh. Organizing violence. *Journal of conflict resolution*, 46(5):599–628, October 2002.

Mary Beard. *SPQR: A History of Ancient Rome*. Liveright, 2015.

Jeanet Sinding Bentzen, Nicolai Kaarsen, and Asger Moll Wingender. Irrigation and autocracy. *Journal of the European Economic Association*, 15(1):1–53, February 2017.

Timothy Besley and Torsten Persson. The origins of state capacity: property rights, taxation and politics. *American Economic Review*, 99(4):1218–1244, 2009.

Lewis R. Binford. *Constructing Frames of Reference: An Analytical Method for Archaeological Theory Building Using Hunter-Gatherer and Environmental Data Sets*. University of California Press, Berkeley, 2001.

Alberto Bisin and Thierry Verdier. The economics of cultural transmission and socialization. In Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, editors, *Handbook of Social Economics*, chapter 9. North Holland, 2011.

Francis Bloch, Garance Genicot, and Debraj Ray. Informal insurance in social networks. *Journal of Economic Theory*, 143:36–58, 2008.

Simon Board. Relational contracts and the value of loyalty. mimeo, UCLA, February 2008.

Robert Boehm. *Hierarchy in the Forest*. Harvard University Press, 1999.

Carles Boix. *Political Order and Inequality*. Cambridge University Press, 2015.

Carles Boix and Frances Rosenbluth. Bones of contention: The political economy of height inequality. *American Political Science Review*, 108(1):1–22, February 2014.

Samuel Bowles and Jung-Kyoo Choi. Coevolution of farming and private property during the early holocene. *Proceedings of the National Academy of Sciences*, 110(22): 8830–3335, May 2013.

Robert Boyd and Peter J. Richerson. The evolution of norms: an anthropological view. *Journal of Theoretical and Institutional Economics*, 150(1):72–87, 1994.

Yann Bramoulle and Sanjeev Goyal. Favoritism. *Journal of Development Economics*, 122:16–27, April 2016.

T. Corey Brennan. *The Praetorship in the Roman Republic*. Oxford University Press, Oxford, 2000. 2 vols.

P. A. Brunt. *Italian Manpower 225 B.C.–A.D. 14*. Oxford University Press, Oxford, 1971.

H. L. Carmichael and W.B. Macleod. Gift giving and the evolution of cooperation. *International Economic Review*, 38:485–508, 1997.

Jean-Paul Carvalho. Veiling. *Quarterly Journal of Economics*, 128(1):336–370, 2013.

Jean-Paul Carvalho. Identity-based organizations. *American Economic Review: Papers and Proceedings*, 106(5):410–414, 2016.

Daniel Clark, Drew Fudenberg, and Alexander Wolitzky. Record-keeping and cooperation in large societies. *Review of Economic Studies*, 88(5):2179–2209, September 2021.

Abner Cohen. *Customs and politics in Urban Africa: a study of Hausa migrants in Yoruba towns*. University of California Press, Berkeley, 1969.

Tapajit Dey, Audris Mockus, and Bogdan Vasilescu. Patterns of effort distribution in open-source software. *Empirical Software Engineering*, 25(4):3777–3814, 2020.

Avinash Dixit. Trade expansion and contract enforcement. *Journal of Political Economy*, 111(6):1293–1317, December 2003.

Matthias Doepke and Fabrizio Zilibotti. Culture, entrepreneurship and growth. working paper 19141, NBER, June 2013.

Gregory K. Dow and Clyde G. Reed. The origins of inequality: Insiders, outsiders, elites and commoners. *Journal of Political Economy*, 121(3):609–641, 2013.

Arthur M. Eckstein. *Rome enters the Greek East : from anarchy to hierarchy in the Hellenistic Mediterranean, 230-170 BC*. Blackwell Pub., Malden, MA, 2008.

Nadia Eghbal. *Working in Public: The Making and Maintenance of Open Source Software*. Stripe Press, 2020.

Glenn Ellison. Cooperation in the prisoner's dilemma with anonymous random matching. *Review of Economic Studies*, 61(3):567–588, July 1994.

Matthias Flückiger, Mario Larch, Markus Ludwig, and Luigi Pascali. The dawn of civilization: Metal trade and the rise of hierarchy. CESifo Working Paper 10929, CESifo, 2024. Forthcoming, *American Economic Review*.

Eric Friedman and Paul Resnick. The social cost of cheap pseudonyms. *Journal of Economics and Management Strategy*, 10(2):173–199, 2001.

Roland G. Fryer Jr. An economic approach to cultural capital. mimeo, University of Chicago, October 2002.

Drew Fudenberg and Jean Tirole. *Game Theory*. MIT Press, Cambridge, MA, 1991.

Philippe Gauthier. Générosité romaine et 'avarice' grecque: sur l'octroi du droit du cité. In *Melanges d'histoire ancienne offerts á William Seston*, pages 207–215. de Boccard, Paris, 1974.

Robert Gibbons. What is economic sociology and should any economists care? *Journal of economic perspectives*, 19(1):3–7, Winter 2005.

Joseph Greenberg. Islam and clan organization among the Hausa. *Southwestern Journal of Anthropology*, 3(3):193–211, Autumn 1947.

Avner Greif. Contract enforceability and economic institutions in early trade: the Maghribi traders' coalition. *American Economic Review*, 83(3):525–547, June 1993.

Avner Greif. Impersonal exchange and the origin of markets: From the community responsibility system to individual legal responsibility in pre-modern Europe. In Masahiko Aoki and Yujiro Hayami, editors, *Communities and Markets in Economic Development*, pages 3–41. Oxford University Press, Oxford, 2001.

Benjamin E. Hermalin. Leadership and corporate culture. In Robert Gibbons and John Roberts, editors, *The Handbook of Organizational Economics*, pages 432–478. Princeton University Press, 2013.

Werner Hildenbrand. *Core and Equilibria of a Large Economy.* Princeton University Press, Princeton, 1974.

Laurence R. Iannaccone. Sacrifice and stigma: reducing free-riding in cults, communes and other collectives. *Journal of Political Economy*, 100(2):271–291, April 1992.

Matthew O. Jackson. A survey of models of network formation: stability and efficiency. In Gabrielle Demange and Myrna Wooders, editors, *Group formation in economics: networks, clubs and coalitions*, Cambridge, UK, August 2003. Cambridge University Press.

Matthew O. Jackson, Tomas Rodriguez-Barraquer, and Xu Tan. Social capital and social quilts: network patterns of favor exchange. *American Economic Review*, 102 (5):1857–1897, August 2012. Cited in manuscript as Jackson:Rodriguez:21; year likely typo for :12.

Michihiro Kandori. Social norms and community enforcement. *Review of Economic Studies*, 59(1):63–80, January 1992.

Lawrence H. Keeley. Hunter-gatherer economic complexity and 'population pressure': A cross-cultural analysis. *Journal of Anthropological Archaeology*, 7(4):373–411, December 1988.

Lamont DeHaven King. *Africa and the Nation State: State formation and identity in ancient Egypt, Hausaland and Southern Africa.* The Edwin Mellen Press, Lewiston, 2006.

Benjamin Klein and Kenneth B. Leffler. The role for market forces in assuring contractual performance. *Journal of Political Economy*, 89(4):615–41, 1981.

Bruce Knauft. Violence and sociality in human evolution [and comments and replies]. *Current Anthropology*, 32(4):391–428, Aug-Oct 1991.

Rachel E. Kranton. The formation of cooperative relationships. *Journal of Economics, Law and Organisation*, 12(1):214–33, April 1996.

David M. Kreps. Corporate culture and economic theory. In James E. Alt and Kenneth A. Shepsle, editors, *Perspectives on Positive Political Economy*, pages 90–143. Cambridge University Press, 1990.

David M. Kreps and Robert Wilson. Sequential equilibria. *Econometrica*, 50(4):863–894, July 1982.

Murray Last. Some economic aspects of conversion in Hausaland. In Nehemia Levitzion, editor, *Conversion to Islam*, chapter 13, pages 236–246. Holmes and Meier, New York, 1979.

Murray Last. History as religion: De-constructing the magians 'maguzawa' of nigerian hausaland. In Jean-Pierre Chrétien, editor, *L'invention religieuse en Afrique*, pages 236–246. Karthala, 1999.

Joshua Lerner and Jean Tirole. The economics of technology sharing: open source and beyond. *Journal of Economic Perspectives*, 19(2):99–120, Spring 2005.

Andrew Lintott. *The Constitution of the Roman Republic*. Oxford University Press, Oxford, 1999.

George Mailath, Andrew Postlewaite, and Larry Samuelson. Buying locally. *International Economic Review*, 57(4):1179–1200, November 2016a.

George J. Mailath and Larry Samuelson. *Repeated Games and Reputations: Long-Run Relationships*. Oxford University Press, Oxford, 2006.

George J. Mailath, Andrew Postlewaite, and Larry Samuelson. Premuneration values and investments in matching markets. *Economic Journal*, 127(604):2041–2065, 2016b. Wiley Online "Early View" 2016; assigned to issue 604 (2017).

Joram Mayshar, Omer Moav, Zvika Neeman, and Luigi Pascali. Cereals, appropriability and hierarchy. mimeo, Warwick, 2015.

Joram Mayshar, Omer Moav, and Luigi Pascali. The origin of the state: land productivity or appropriability? *Journal of Political Economy*, 130(4):1091–1144, April 2022.

John McMillan. *Reinventing the Bazaar: a natural history of markets*. W.W. Norton & Co., New York, 2002.

Stelios Michalopoulos and Elias Papaioannou. Historical legacies and African development. *Journal of Economic Literature*, 58(1):53–128, March 2020.

Claude Nicolet. *The World of the Citizen in Republican Rome*. University of California Press, Berkeley, 1980. Translated by P. S. Falla from *Le métier de citoyen dans la Rome républicaine* (Paris: Gallimard, 1976).

Douglass North, John Wallis, and Barry Weingast. *Violence and social orders: a conceptual framework for interpreting recorded human history*. Cambridge University Press, 2009.

Martin A. Nowak and Karl Sigmund. Evolution of indirect reciprocity. *Nature*, 437: 1291–1298, October 2005.

Josiah Ober. *The Rise and Fall of Classical Greece*. Princeton University Press, 2015.

Wojciech Olszewski and Mikhail Safronov. Efficient cooperation by exchanging favors. *Theoretical Economics*, 13(3):1191–1231, 2018a.

Wojciech Olszewski and Mikhail Safronov. Efficient chip strategies in repeated games. *Theoretical Economics*, 13:951–977, 2018b.

T. Douglas Price and James A. Brown. *Prehistoric Hunter-Gatherers: The Emergence of Cultural Complexity*. Academic Press, Orlando, 1985.

Garey Ramey and Joel Watson. Bilateral trade and opportunism in a matching market. *Contributions to theoretical economics*, 1(1), 2001.

Eric S. Raymond. *The Cathedral and Bazaar: musings on linux and open source by an accidental revolutionary*. O' Reilly, Cambridge, MA, 2001.

Frank A. Salamone. *The Hausa of Nigeria*. The University Press of America, 2010.

Paul Seabright. *The Birth of Hierarchy.*, chapter 5, pages 109–117. In Sterelny et al. [2013], 2013.

Sonali Shah. Motivation, governance and the viability of hybrid forms in open source software development. *Management Science*, 52(7):1000–1014, 2006.

Carl Shapiro and Joseph E. Stiglitz. Equilibrium unemployment as a worker discipline device. *American Economic Review*, 74(3):433–444, 1984.

A. N. Sherwin-White. *The Roman Citizenship*. Oxford University Press, Oxford, 2nd edition, 1973.

Joel Sobel. For better or forever: formal versus informal enforcement. *Journal of Labor Economics*, 2006.

Kim Sterelny. *Life in Interesting Times: Cooperation and Collective Action in the Holocene*. In Sterelny et al. [2013], 2013.

Kim Sterelny, Richard Joyce, Brett Calcott, and Ben Fraser, editors. *Cooperation and its evolution*. MIT Press, Cambridge, MA, 2013.

Nancy L. Stokey, Robert E. Lucas, and Edward C. Prescott. *Recursive Methods in Economic Dynamics*. Harvard University Press, Cambridge, MA, 1989.

Jean Tirole. Digital dystopia. *American Economic Review*, 111(6):2007–2048, 2021.

Eric Van den Steen. On the origin of shared beliefs (and corporate culture). MIT Sloan working paper 4553-05, August 2005.

Harrison C. White. *Chains of opportunity: system models of mobility in organizations*. Harvard University Press, Cambridge, MA, 1970.

Alexander Wolitzky. Communication with tokens in repeated games on networks. *Theoretical Economics*, 10:67–101, 2015.

James Woodburn. Egalitarian societies. *Man*, 17(3):431–451, September 1982.

Norman Yoffee. *Myths of the Archaic State: Evolution of the Earliest Cities, States and Civilizations*. Cambridge University Press, Cambridge, 2005.