Experimentation in Networks

Simon Board and Moritz Meyer-ter-Vehn

C Online Appendix

C.1 Proof of Strategic Substitutes

Here we show that i's social learning $\{\beta_{i,t}\}$ rises in other agents' cutoffs τ_{-i}. Assume others raise their cutoffs, $\tau_{-i} \leq \tau'_{-i}$. Realize independent "potential success times" $\{T^{\iota}_{-i}\}$ according to independent Poisson processes with arrival rate 1; j's first actual success time T_j equals $\inf\{T^{\iota}_j : A_{j,T^{\iota}_j} = 1\}$. Write T_{-i}, T'_{-i} for the first actual success times of $j \neq i$ given cutoffs $\tau_{-i} \leq \tau'_{-i}$, assuming throughout that no agent $j \neq i$ ever observes a success by i.

Since we have fixed strategies, each agent $j \neq i$ succeeds earlier when they use the higher cutoffs, $T_j \geq T'_j$ for all $\{\tilde{T}^{\iota}_{-i}\}$. This follows by induction over the contagion process with initial successes during experimentation as induction anchor. Hence agent i sees a success earlier, $S_i = \min_{j\in N_i(G)} T_j \geq \min_{j\in N_i(G')} T'_j = S'_i$ for all $\{\tilde{T}^{\iota}_{-i}\}$, and so i's social learning curve is higher $\beta_{i,t} = -\log \Pr^{-i}(t < S_i) \leq -\log \Pr^{-i}(t < S'_i) = \beta'_{i,t}$. By Lemma 1, the associated cutoffs are ranked $\tau_i \geq \tau'_i$, so cutoffs are strategic substitutes.

This proof also implies that social learning $\{\beta_{i,t}\}$ increases in network density for fixed τ_{-i}. Specifically, order deterministic networks by set inclusion $g \subseteq g'$ in $\{1, ..., I\}^2$, and extend this order to random networks, by writing $G \preceq G'$ if they are coupled to networks $\tilde{G} \subseteq \tilde{G}'$.[40] Then, we get $T_j \geq T'_j$ also for all realizations of $\{T^{\iota}_{-i}\}$ and the coupled networks $\tilde{G}, \tilde{G}'$. Thus, a rise in G raises $\{\beta_{i,t}\}$ and lowers τ_i by Lemma 1.

C.2 Proof of Proposition 3 (Equal Cutoffs of Equals)

The result uses two Lemmas. For social learning $\{\beta_t\}$ and the associated optimal cutoff τ, define total learning $\beta_t + \min\{t, \tau\}$. So defined, $\Pr^H(\min\{S, T\} \leq t) = 1 - \exp(-(\beta_t + \min\{t, \tau\}))$, where $\Pr^H$ is taken over the network G and success times of all agents $\{T_j\}$ including i, conditional on $\theta = H$.

Lemma 8. *Higher total learning, $\beta_t + \min\{\tau, t\} \geq \hat{\beta}_t + \min\{\hat{\tau}, t\}$ for all t, is associated with a lower optimal cutoff, $\tau \leq \hat{\tau}$.*

[40] Random variables X, X' are coupled to $\tilde{X}, \tilde{X}'$ if they have the same marginal distributions and $\tilde{X}, \tilde{X}'$ are defined on the same probability space.

This is closely related to Lemma 1, that lower social learning $\{\beta_t\} \leq \{\hat{\beta}_t\}$ implies higher cutoffs $\tau \geq \hat{\tau}$. Lemma 8 shows additionally that the higher cutoff cannot lead to higher total learning. Intuitively, all learning (both social and own) crowds out incentives.

Lemma 9. *Fix a network G, cutoffs $\{\tau_k\}_{k\neq i,j}$ and $\tau_* < \tau^*$, and write k's first success time as $\{T_k\}$ if $\tau_i = \tau^*, \tau_j = \tau_*$, and $\{T'_k\}$ if $\tau_i = \tau_*, \tau_j = \tau^*$. Then* $\min\{T_i, S_i\} \overset{D}{\preceq} \min\{T'_i, S'_i\}$.[41]

Lemma 9 is intuitive: Additional experimentation during $[\tau_*, \tau^*]$ is more immediate and useful to i when done by i herself instead of j.

Proof of Proposition 3. By contradiction, assume $\tau_i > \tau_j$. Exchangeability, $G_{i\leftrightarrow j} = G$, implies $\min\{T_j, S_j\} \overset{D}{=} \min\{T'_i, S'_i\}$. Lemma 9 then implies $\min\{T_i, S_i\} \overset{D}{\preceq} \min\{T_j, S_j\}$. Noting the connection between total learning and the time of the first observed success, $\Pr^H(\min\{S, T\} \leq t) = 1 - \exp(-(\beta_t + \min\{\tau, t\}))$, this implies $\{\beta_{i,t} + \min\{\tau_i, t\}\} \geq \{\beta_{j,t} + \min\{\tau_j, t\}\}$ and so, by Lemma 8, $\tau_i \leq \tau_j$. □

Proof of Lemma 8. Lemmas 1 and 6 study incentives ψ_τ as a function of social learning $\{\beta_t\}$; we now study ψ_τ as a function of total learning $\{\beta_t + \min\{t, \tau\}\}$.

By contradiction assume that $\beta_t + \min\{\tau, t\} \geq \hat{\beta}_t + \min\{\hat{\tau}, t\}$ for all t, yet $\tau > \hat{\tau}$. Define $\tilde{\beta}_t := \hat{\beta}_t - (\tau - \hat{\tau})$; clearly $\tilde{\beta}_t \leq \beta_t$, and so Lemma 1 implies

$$\psi_\tau(\{\tilde{\beta}_t\}) \geq \psi_\tau(\{\beta_t\}) = 0.$$

Since $\tilde{\beta}_\tau + \tau = \hat{\beta}_\tau + \hat{\tau}$ and $\tilde{b}_u = \hat{b}_u$ for $u \geq \tau$, time-τ experimentation incentives for the social learning curve $\{\hat{\beta}_t\}$ are also positive

$$e^{\int_0^\tau r + p_u(\hat{a}_u + \hat{b}_u)du} \frac{\partial \hat{\Pi}(\mathbb{I}_{\{t\leq\hat{\tau}\}})}{\partial a_\tau} = P^\emptyset(\hat{\beta}_\tau + \hat{\tau}) \left(x + ry \int_\tau^\infty e^{-\int_\tau^s (r + \hat{b}_u)du} ds \right) - c = \psi_\tau(\{\tilde{\beta}_t\}) \geq 0$$

where the first equality follows as in (27), using $\hat{a}_u = 0$ at $u \geq \tau$ since $\tau > \hat{\tau}$. Front-loading, (22), then implies

$$\frac{\partial \hat{\Pi}(\mathbb{I}_{\{t\leq\hat{\tau}\}})}{\partial a_{\hat{\tau}}} > \frac{\partial \hat{\Pi}(\mathbb{I}_{\{t\leq\hat{\tau}\}})}{\partial a_\tau} \geq 0$$

contradicting the optimality of cutoff $\hat{\tau}$. □

Proof of Lemma 9. We couple $\min\{T_i, S_i\}$ and $\min\{T'_i, S'_i\}$ by first realizing any agent k's first success time $\bar{T}_k$ when i and j both use cutoff τ_* in network G. When we raise i's cutoff to $\tau_i = \tau^*$ while keeping $\tau_j = \tau_*$, we account for i's potential additional experimentation

[41] As always, $S_i = \min_{j\in N_i(G)}\{T_j\}$ and $S'_i = \min_{j\in N_i(G)}\{T'_j\}$.

over $[\tau_*, \tau^*]$ by realizing an exponential random variable $Z \sim Exp(1)$ and setting $\bar{Z} = Z$ if $Z \leq \tau^* - \tau_*$ and $\bar{Z} = \infty$ otherwise. So constructed, $\min\{T_i, S_i\} \stackrel{D}{=} \min\{\bar{T}_i, \bar{S}_i, \tau_* + \bar{Z}\}$.

Analogously when we raise j's cutoff to $\tau_j = \tau^*$ while keeping $\tau_i = \tau_*$, we account for j's potential additional experimentation over $[\tau_*, \tau^*]$ with the same random variable $\bar{Z}$. We then realize $\min\{T_i', S_i'\}$ by tracing the additional successes and experimentation through the network. Since this cascade does not start before $\tau_* + \bar{Z}$, we obtain $\min\{T_i', S_i'\} \stackrel{D}{\succeq} \min\{\bar{T}_i, \bar{S}_i, \tau_* + \bar{Z}\} \stackrel{D}{=} \min\{T_i, S_i\}$. □

C.3 Proof of Lemma 3 (Links in Large Random Networks)

Part (a): We will show separately that for every $\epsilon > 0$

$$\Pr\left[N^I \geq (1+\epsilon) I(1 - e^{-\hat{n}^I/I})\right] \to 0, \tag{43}$$

$$\Pr\left[N^I \leq (1-\epsilon) I(1 - e^{-\hat{n}^I/I})\right] \to 0. \tag{44}$$

This implies that N^I converges to $I(1 - e^{-\hat{n}^I/I})$ in distribution, $N^I/(I(1 - e^{-\hat{n}^I/I})) \stackrel{D}{\to} 1$.

Start with the upper bound, (43). We can restrict attention to $\hat{\rho} = \lim \hat{n}^I/I < \infty$; for $\hat{\rho} = \infty$, we have $(1+\epsilon)I(1 - e^{-\hat{n}^I/I}) > I$ for any $\epsilon > 0$ and large enough I, so trivially $\Pr[N^I \geq (1+\epsilon)I(1 - e^{-\hat{n}^I/I})] = 0$.

Realize Iris's $\hat{n}^I$ stubs k one after another, and keep track of the number of stubs $K^I(m)$ used to reach degree m; if i has less than m neighbors in the realized network set $K^I(m) := \hat{n}^I + 1$. When connecting Iris's k^{th} stub to her m^{th} neighbor, $I - m$ potential new neighbors with $\hat{n}^I(I-m)$ stubs compete with $\hat{n}^I m - (2k-1)$ remaining stubs of Iris and her $m-1$ neighbors, sandwiching the success rate between $\frac{I-m}{I}$ and $\frac{I-m}{I-2}$. Writing Y_ℓ^I for independent (shifted) geometric random variables with success rate $\frac{I-\ell}{I}$ we can thus upper-bound $K^I(m) \stackrel{D}{\preceq} \sum_{\ell=1}^m Y_\ell^I$.

The chance of m or more neighbors is then upper-bounded by

$$\Pr\left[N^I \geq m\right] = \Pr\left[K^I(m) \leq \hat{n}^I\right] \leq \Pr\left[\sum_{\ell=1}^m Y_\ell^I \leq \hat{n}^I\right] \leq \inf_{\xi \geq 0} \exp\left(\xi \hat{n}^I + \sum_{\ell=1}^m \log E[e^{-\xi Y_\ell^I}]\right)$$

$$= \inf_{\xi \geq 0} \exp\left(\xi(\hat{n}^I - m) - \sum_{\ell=1}^m \log \frac{1 - e^{-\xi}\ell/I}{1 - \ell/I}\right) \tag{45}$$

where the second inequality is a Chernoff-bound, and the final equality evaluates the moment generating function of the shifted geometric distribution, $E[e^{-\xi Y_\ell^I}] = \frac{e^{-\xi}(1-\ell/I)}{1-e^{-\xi}\ell/I}$.

Since $\log \frac{I-e^{-\xi}\ell}{I-\ell}$ rises in ℓ, the last term in (45) is lower-bounded by

$$\sum_{\ell=1}^{m} \log \frac{1-e^{-\xi}\ell/I}{1-\ell/I} \geq \int_0^m \left(\int_{1-\ell/I}^{1-e^{-\xi}\ell/I} \frac{1}{x} dx \right) d\ell = \int_{1-m/I}^{1} \left(\int_{I(1-x)}^{\min\{e^{\xi} I(1-x), m\}} \frac{1}{x} d\ell \right) dx$$
$$= \int_{1-m/I}^{1-e^{-\xi}m/I} \frac{m - I(1-x)}{x} dx + \int_{1-e^{-\xi}m/I}^{1} \frac{I(1-x)(e^{\xi}-1)}{x} dx$$
$$= I\left[(1-m/I)\log(1-m/I) - e^{\xi}\left(1-e^{-\xi}m/I\right)\log(1-e^{-\xi}m/I)\right].$$

For any $\epsilon > 0$, we now set $m = m^I := \left\lceil (1+\epsilon) I (1-e^{-\hat{n}^I/I}) \right\rceil$, substitute back into the term in parentheses in (45), and divide by I

$$\xi \frac{\hat{n}^I - m^I}{I} - (1-m^I/I)\log(1-m^I/I) + e^{\xi}\left(1-e^{-\xi}m^I/I\right)\log(1-e^{-\xi}m^I/I) =: \Gamma^I(\xi,\epsilon)$$

with limit $\Gamma(\xi,\epsilon)$ as $I \to \infty$. So defined, (45) becomes

$$\Pr\left[N^I \geq (1+\epsilon) I (1-e^{-\hat{n}^I/I})\right] \leq \inf_{\xi \geq 0} \exp\left(I \Gamma^I(\xi,\epsilon)\right) \tag{46}$$

The derivative $\Gamma_\xi(0,\epsilon) = \hat{\rho} + \log(1-(1+\epsilon)(1-e^{-\hat{\rho}}))$ vanishes for $\epsilon = 0$ and falls in ϵ. Thus, for any $\epsilon > 0$ we have $\Gamma_\xi(0,\epsilon) < 0$. Also, $\Gamma(0,\epsilon) = 0$, and so $\Gamma(\xi,\epsilon) < 0$ for small ξ, and $\Gamma^I(\xi,\epsilon)$ is boundedly negative for large I. Thus, the RHS of (46) vanishes for $I \to \infty$, implying (43).

The lower bound (44) follows analogously.

Part (b): Since $N^I/I \leq 1$, part (a) implies convergence in expectation $\lim n^I/I = \lim(1 - e^{-\hat{n}^I/I}) = 1 - e^{-\hat{\rho}}$. Further, since

$$I(1-\exp^{-\hat{n}^I/I}) = I\left(\frac{\hat{n}^I}{I} - \frac{1}{2}\left(\frac{\hat{n}^I}{I}\right)^2 + \frac{1}{6}\left(\frac{\hat{n}^I}{I}\right)^3 - ...\right)$$

$I(1-\exp^{-\hat{n}^I/I})/\hat{n}^I - 1$ (and hence $n^I/\hat{n}^I - 1$) is of order $\hat{n}^I/I$, which vanishes for $\hat{\rho} = 0$.

Part (c): Since $A_t^I = 1$ for $t < \tau^I$, we have $\beta_{\tau^I}^I = \int_0^{\tau^I} E^{-i}[N^I | t < S_i^I] dt$. For I finite, $E^{-i}[N^I | t < S_i^I] < n^I$ (and so $\beta_{\tau^I}^I < n^I \tau^I$) because lack of success, $t < S_i^I$, indicates fewer neighbors N^I. To bound the effect of such updating, we note that conditional on $|N^I - n^I| \leq$

ϵn^I, and so $N^I \leq (1+\epsilon)n^I$, we have $\Pr^{-i}(t < S_i^I || N^I - n^I| \leq \epsilon n^I) \geq e^{-(1+\epsilon)n^I t}$. Thus

$$\frac{\Pr^{-i}(|N^I - n^I| \leq \epsilon n^I | t < S_i^I)}{\Pr^{-i}(|N^I - n^I| > \epsilon n^I | t < S_i^I)} = \frac{\Pr^{-i}(|N^I - n^I| \leq \epsilon n^I)}{\Pr^{-i}(|N^I - n^I| \geq \epsilon n^I)} \frac{\Pr^{-i}(t < S_i^I || N^I - n^I| \leq \epsilon n^I)}{\Pr^{-i}(t < S_i^I || N^I - n^I| > \epsilon n^I)}$$

$$\geq \frac{\Pr^{-i}(|N^I - n^I| \leq \epsilon n^I)}{\Pr^{-i}(|N^I - n^I| \geq \epsilon n^I)} e^{-(1+\epsilon)n^I t} \tag{47}$$

We show below that $n^I \tau^I$ is bounded. This bounds $e^{-(1+\epsilon)n^I t}$ away from 0 for all $t \leq \tau^I$. Thus, as the prior likelihood-ratio of $|N^I - n^I| \leq \epsilon n^I$ on the RHS of (47) diverges as $I \to \infty$ (by part (a)), so does the posterior likelihood-ratio on the LHS of (47). By part (b), N^I/n^I is bounded, implying $E^{-i}[N^I | t < S_i^I]/n^I \to 1$ and so $\beta_{\tau^I}^I/(n^I \tau^I) \to 1$, finishing the proof of part (c).

To show that $n^I \tau^I$ is bounded, assume it was not. Then we could choose $\hat{\tau}^I < \tau^I$ such that $n^I \hat{\tau}^I$ is bounded, but with limit $\lim n^I \hat{\tau}^I > \bar{\tau}$. Applying the above argument to $n^I \hat{\tau}^I$ instead of $n^I \tau^I$, we get $\lim \beta_{\hat{\tau}^I}^I = \lim n^I \hat{\tau}^I > \bar{\tau}$, and so $p_{\tau^I} < p_{\hat{\tau}^I} = P^\emptyset(\beta_{\hat{\tau}^I}^I + \hat{\tau}^I) < \underline{p}$ for large I, contradicting $p_\tau \in [\underline{p}, \bar{p}]$ as illustrated in Figure 1.

C.4 Proof of Equation (18)

We apply Bayes' rule

$$1 - a_t = \frac{\Pr^{-\ell}(\forall k, \ell' : t < T_k, T_{\ell'})}{\Pr^{-\ell}(t < T_\ell, \forall k : t < T_k)} = \begin{cases} \frac{\exp(-(K+L)t)}{\exp(-(K+1)t)} = \exp\left(-(L-1)t\right) & t < \tau_k \\ \frac{\exp(-K\tau_k - Lt))}{\exp\left(-K\left(\tau_k + \int_{\tau_k}^t a_s ds\right) - t\right)} = \exp\left(-(L-1)t + K\int_{\tau_k}^t a_s ds\right) & t \in (\tau_k, \tau_\ell) \\ \frac{\exp(-K\tau_k - L\tau_\ell))}{\exp\left(-K\left(\tau_k + \int_{\tau_k}^t a_s ds\right) - \tau_\ell\right)} = \exp\left(-(L-1)\tau_\ell + K\int_{\tau_k}^t a_s ds\right) & t > \tau_\ell \end{cases}$$

and then differentiate wrt t.

C.5 Proof of Theorem 2 (Core-Periphery Networks)

The challenge with this proof is the complexity of characterizing two outcome variables, asymptotic information and welfare, for a myriad of cases. Specifically we must consider six different network densities $\kappa \lesseqgtr \kappa^*$, $\rho = 0$, $\in (0,1)$, or $= 1$, and pessimistic priors $p_0 < \bar{p}$ as well as optimistic ones. While some arguments apply to all of these cases, each case also has its idiosyncrasies.

We structure the exposition in order of increasing network density, characterizing asymptotic information and welfare in parallel and emphasizing the case of pessimistic priors $p_0 < \bar{p}$. But to avoid repetitions, we sometimes break this linear narrative by bracketing out arguments that apply more broadly.

As in the paper body, we superscript variables in finite networks with the network size I, e.g. τ_ℓ^I, and drop the superscript in the limit, e.g. $\tau_\ell := \lim_{I\to\infty} \tau_\ell^I$. A priori the limit is well-defined only for some subsequence, but the analysis characterizes all limits under consideration uniquely.

Asymptotic information equals $\beta = \lim \beta^I = \lim(K^I\tau_k^I + L^I\tau_\ell^I)$ since the network is connected and each agent's own experimentation $\tau_{k,\ell}^I$ (which in principle is excluded from the social information β) is negligible as $I \to \infty$. It will be useful to decompose β into core agents' pre-, and post-cutoff learning

$$\Upsilon_k^I := I\tau_k^I \qquad \Upsilon_\ell^I := L^I(\tau_\ell^I - \tau_k^I).$$

We can already note two bounds on $\Upsilon_k, \Upsilon_\ell$: Total information $\beta = \Upsilon_k + \Upsilon_\ell$ is strictly positive: By contradiction, $\beta = 0$ means agents face the single-agent problem, choose $\tau_k = \tau_\ell = \bar{\tau} > 0$ and so $\beta = \infty$. Any agent's pre-cutoff learning β_τ is no larger than $\bar{\tau}$, recalling from (4) that $P^\emptyset(\beta_\tau)(x+y) - c \geq \psi_\tau = 0$. For core agents, this means $\Upsilon_k \leq \bar{\tau}$. Thus, there is asymptotic learning iff $\Upsilon_\ell = \infty$; a sufficient (but not necessary) condition is $\tau_\ell > 0$.

C.5.1 Case 1: Bounded core size $\kappa < \infty$

Preliminaries. We first establish a necessary and sufficient condition for maximal social learning by peripherals

$$\beta_{\ell,t} \equiv \kappa t \qquad \text{iff} \qquad \Upsilon_\ell = \infty. \tag{48}$$

If $\Upsilon_\ell = \infty$, core agents immediately observe a peripheral succeed, and then work forever after. If $\Upsilon_\ell < \infty$, the probability of a success $1 - e^{-(\Upsilon_k + \Upsilon_\ell)}$ is less than one, bounding above $b_{\ell,t} \leq \kappa(1 - e^{-(\Upsilon_k+\Upsilon_\ell)}) < \kappa$ for $t > \tau_k$.

By Lemma 1, the social learning upper-bound (48) implies an incentive lower-bound

$$\psi_{\ell,0} \geq \underline{\psi}_{\ell,0}^\kappa := p_0\left(x + \frac{r}{r+\kappa}y\right) - c \tag{49}$$

with equality iff $\Upsilon_\ell = \infty$.

We distinguish three cases, $\kappa \lesseqgtr \kappa^*$; for optimistic priors $p_0 \geq \bar{p}$, we have $\kappa^* = \infty$, and so only case 1a is relevant.

Case 1a: $\kappa < \kappa^*$**.** Since $\underline{\psi}_{\ell,0}^\kappa$ falls in κ, we have $\underline{\psi}_{\ell,0}^\kappa > \underline{\psi}_{\ell,0}^{\kappa^*} = 0$, so $\psi_{\ell,0} > 0$, and continuity of $\psi_{\ell,0}$ implies $\tau_\ell > 0$, and asymptotic learning $\Upsilon_\ell = \infty$. By Lemma 2, welfare is bounded below the benchmark $\mathcal{V}(\tau_\ell, \kappa\tau_\ell) < \mathcal{V}(0,0) = V^*$. Quantitatively, $\Upsilon_\ell = \infty$ and (48) imply

$\beta_{\ell,t} = \kappa t$, so welfare increases in κ by Lemma 1.

For $p_0 \geq \bar{p}$, only one argument needs adapting: the welfare benchmark now equals $V^* = p_0 y$ which requires immediate and perfect social learning, $\beta_t = \infty$ for $t > 0$. Clearly, $\beta_{\ell,t} = \kappa t$ falls short of this benchmark.

Case 1b: $\kappa = \kappa^*$**.** Now $\underline{\psi}^{\kappa}_{\ell,0} = 0$. We show asymptotic learning, $\Upsilon_\ell = \infty$, by contradiction: By (49), $\Upsilon_\ell < \infty$ would imply $\psi_{\ell,0} > 0$ and so $\tau_\ell > 0$, leading to the contradiction that $\Upsilon_\ell = \infty$. In turn, $\Upsilon_\ell = \infty$ implies by (48) and (49) that $\psi_{\ell,0} = \underline{\psi}^{\kappa}_{\ell,0} = 0$ and so $\tau_\ell = 0$ and $\kappa\tau_\ell = 0$, attaining the welfare benchmark $\mathcal{V}(0,0) = V^*$.

Case 1c: $\kappa \in (\kappa^*, \infty)$**.** Now $\underline{\psi}^{\kappa}_{\ell,0} < 0$. Asymptotic learning fails because $\Upsilon_\ell = \infty$ would imply by (48) and (49) that $\psi_{\ell,0} = \underline{\psi}^{\kappa}_{\ell,0} < 0$ and so $\tau^I_\ell = 0$ for large I and $\Upsilon_\ell = 0$. In turn, $\Upsilon_\ell < \infty$ implies $\tau_\ell = 0$ and $\psi_{\ell,0} = 0$. To quantify information, we first claim that $\Upsilon_k = \lim I\tau^I_k = 0$: Indeed, core agents receive all social information immediately, $\beta_{k,t} = \Upsilon_k + \Upsilon_\ell$ for all $t > 0$, while peripherals' learning is bounded by $\beta_{\ell,t} \leq \kappa t$. This bounds incentives of core agents above $\psi_{k,0} < \psi_{\ell,0} = 0$, and so $\tau^I_k = 0$ for large I.[42]

Social information thus equals Υ_ℓ. We now show this falls in κ: Peripherals observe a success by time t iff at least one peripheral succeeds during experimentation, and then a core agent succeeds during $(0,t]$; thus $1 - e^{-\beta_{\ell,t}} = (1 - e^{-\Upsilon_\ell})(1 - e^{-\kappa t})$.[43] Since the RHS rises with both κ and Υ_ℓ and experimentation incentives $\psi_{\ell,0}$ fall in $\{\beta_{\ell,t}\}$, the equilibrium condition $\psi_{\ell,0} = 0$ implies that a rise in information transmission κ must be compensated by a fall in aggregate information Υ_ℓ. For future reference, we note that as $\kappa \to \infty$, the learning curve $\beta_{\ell,t}$ converges to Υ_ℓ for each $t > 0$, and so peripherals' indifference condition converges to $p_0(x + e^{-\Upsilon_\ell} y) = c$, pinning down aggregate information Υ_ℓ.

Finally, since $\tau_\ell = \kappa\tau_\ell = 0$, welfare attains the benchmark $\mathcal{V}(0,0) = V^*$.

C.5.2 Case 2: Exploding core $\kappa = \infty$

Preliminaries. We first assume $\rho < 1$, and cover the case $\rho = 1$ separately. We prepare the ground with two preliminary lemmas.

Lemma 10. *Assume $\kappa = \infty$, $\rho < 1$, and any prior $p_0 > \underline{p}$.*

(*a*) *Individual learning vanishes: $\tau^I_k, \tau^I_\ell \to 0$.*

(*b*) *Social learning is immediate: For all $t > 0$, $\beta^I_{k,t}, \beta^I_{\ell,t} \to \Upsilon_k + \Upsilon_\ell$.*

[42] We also get $\tau^I_k = 0$ for large I and $\Upsilon_k = 0$ in cases 1a,b with $p_0 < \bar{p}$, where $\psi_{k,0} < 0$ is ensured by $\beta_{k,t} = \infty$ for all $t > 0$.

[43] Solving for $\beta_{\ell,t}$ and differentiating yields $b_{\ell,t} = \kappa \frac{e^{-\kappa t}(1-e^{-\Upsilon_\ell})}{e^{-\kappa t}(1-e^{-\Upsilon_\ell})+e^{-\Upsilon_\ell}}$, generalizing (48).

Proof. Part (a) follows by the upper bound on pre-cutoff learning $\beta_\tau \le \bar{\tau}$. For core agents, $\beta^I_{k,\tau^I_k} = (I-1)\tau^I_k \le \bar{\tau}$. For peripherals,

$$\beta^I_{\ell,\tau^I_\ell} = K^I \tau^I_k + \int_{\tau^I_k}^{\tau^I_\ell} K^I a^I_t dt \tag{50}$$

where core agents' expected effort a^I_t from (18) drifts towards $\min\{(L^I-1)/K^I, 1\}$ and is hence bounded away from 0 by our assumption that $\rho < 1$. The upper bound, $\beta^I_{\ell,\tau_\ell} < \bar{\tau}$ thus requires the domain to vanish, $\tau^I_\ell \to 0$, as the integrand explodes, $K^I \to \infty$.

Turning to part (b), the conditional probability that some agent i has observed a neighbor succeed by $t > \tau^I_\ell$ is sandwiched via

$$\left(1-\exp(-(I\tau^I_k + (L^I-1)(\tau^I_\ell - \tau^I_k)))\right)\left(1-\exp(-K^I(t-\tau^I_\ell))\right) < 1-\exp(-\beta^I_{i,t}) < 1-\exp(-(\Upsilon^I_k+\Upsilon^I_\ell))$$

The upper bound is the probability that any agent succeeds. The lower bound is the probability that some agent $j \neq i$ succeeds during experimentation, times the probability that a core agent succeeds in $[\tau^I_\ell, t]$. Both bounds converge to $1-\exp(-(\Upsilon_k+\Upsilon_\ell))$ as $I \to \infty$. □

Lemma 10(b) implies that success is observed either immediately, with probability $p_0(1-e^{-(\Upsilon_k+\Upsilon_\ell)})$, or never; so welfare of both core agents and peripherals equals $V_k = V_\ell = p_0(1-e^{-(\Upsilon_k+\Upsilon_\ell)})y$ and our monotonicity results for social information apply equally to welfare.

Lemma 10(b) implies that social learning of both core agents and peripherals occurs in two bursts: one before the cutoff and one immediately after, and both approaching $t=0$. For such learning with burst sizes β^- and β^+, the indifference condition $\psi_t = 0$ becomes

$$\Psi(\beta^-, \beta^+) := P^\emptyset(\beta^-)(x + e^{-\beta^+}y) - c = 0. \tag{51}$$

Recalling the effects of social learning on experimentation incentives (5) and $ry = x - c$, the solution of (51) has slope

$$-\frac{d\beta^+}{d\beta^-} = \frac{\partial_{\beta^-}\Psi}{\partial_{\beta^+}\Psi} = \frac{e^{-\beta^+}y + x - c}{e^{-\beta^+}y} = 1 + re^{\beta^+}. \tag{52}$$

To apply (51) to core agents and peripherals, write asymptotic pre-cutoff learning as $\beta_{\ell,\tau_\ell} = \lim \beta^I_{\ell,\tau^I_\ell}$, experimentation incentives as $\psi_{\ell,\tau_\ell} = \lim \psi^I_{\ell,\tau^I_\ell}$, and similarly for core agents, substituting "k" for "ℓ".[44] For core agents, $\beta^- = \beta_{k,\tau_k} = \Upsilon_k$, $\beta^+ = \Upsilon_\ell$, and (51) coincides with the limit of (17) as $L \to \infty$.

[44]Note that even though $\tau^I_\ell \to \tau_\ell = 0$, β_{ℓ,τ_ℓ} is distinct from, and generally greater than the other limit $\beta_{\ell,0} = \lim \beta^I_{\ell,0} = 0$.

Lemma 11. *Assume $\kappa = \infty$, $\rho < 1$, and any prior $p_0 > \underline{p}$.*

(a) Core agents' indifference condition converges to

$$\Psi(\Upsilon_k, \Upsilon_\ell) = P^\emptyset(\Upsilon_k)\left(x + e^{-\Upsilon_\ell} y\right) - c = 0. \tag{53}$$

(b) Pre-cutoff learning of core agents and peripherals coincides: $\beta_{\ell,\tau_\ell} = \Upsilon_k$.

Proof. Part (a): We will show that core agents' cutoff incentives $\psi^I_{k,\tau^I_k} \to 0$, implying (53). By contradiction, assume without loss that $\lim \psi^I_{k,\tau^I_k} = \psi_{k,0} = p_0\left(x + e^{-\Upsilon_\ell} y\right) - c < 0$. Using Lemma 10(b) (immediate learning by both core agents and peripherals) and the greater importance of pre-cutoff learning (52), strict shirking incentives by core agents carry over to peripherals[45]

$$\psi_{\ell,\tau_\ell} = \Psi(\beta_{\ell,\tau_\ell}, \Upsilon_\ell - \beta_{\ell,\tau_\ell}) \le \Psi(0, \Upsilon_\ell) = \psi_{k,\tau_k} < 0.$$

Thus $\tau^I_\ell = 0$ for large I, so $\Upsilon_k + \Upsilon_\ell = 0$, leading to the contradiction that $\psi_{k,0} = \psi_{\ell,0} = p_0\,(x + y) - c = \Psi(0,0) > 0$.

Part (b): This follows from the fact that core agents and peripherals have the same value, and so $\mathcal{V}(0, \Upsilon_k) = V_k = V_\ell = \mathcal{V}(0, \beta_{\ell,\tau_\ell})$. □

Lemma 11 establishes two conditions for $\Upsilon_k, \Upsilon_\ell$. Below we show they admit a unique solution; a corner solution for $\rho = 0$, and an internal one for $\rho \in (0,1)$.

Case 2a: $\rho = 0$. In this case we get a corner solution for $\Upsilon_k, \Upsilon_\ell$ with $\Upsilon_k/\Upsilon_\ell = 0$. Indeed, using Lemma 11(b), pre-cutoff learning is a vanishing proportion of post-cutoff learning

$$\Upsilon_k = \beta_{\ell,\tau_\ell} = \lim \beta^I_{\ell,\tau^I_\ell} \le \lim K^I \tau^I_\ell = \lim \frac{K^I}{L^I} L^I \tau^I_\ell \le \text{“}\frac{\rho}{1-\rho}(\Upsilon_k + \Upsilon_\ell)\text{”}, \tag{54}$$

where the last inequality is only well-defined if $\Upsilon_\ell < \infty$, and should otherwise be omitted. Since $\rho = 0$, we must have either $\Upsilon_k = 0$ or $\Upsilon_\ell = \infty$ or both.

For pessimistic priors $p_0 < \bar{p}$, core agents' indifference (53) rules out asymptotic learning, so $\Upsilon_\ell < \infty$ and (54) implies $\Upsilon_k = 0$. In turn, aggregate information Υ_ℓ solves $\Psi(0, \Upsilon_\ell) = p_0\left(x + e^{-\Upsilon_\ell} y\right) - c = 0$.[46]

For $p_0 \ge \bar{p}$, Υ_k solves $P^\emptyset(\Upsilon_k) = \bar{p}$ and $\Upsilon_\ell = \infty$.[47] Core agents' indifference (53) clearly requires experimentation until the myopic threshold, $P^\emptyset(\Upsilon_k) \le \bar{p}$. If, by contradiction, core agents experiment past the myopic threshold, $P^\emptyset(\Upsilon_k) < \bar{p}$, then (53) implies $\Upsilon_\ell < \infty$, and

[45] Note the contrast to the case with bounded core size $\kappa < \infty$ (and $p_0 < \bar{p}$), where peripherals learn slower than core agents, so that $\psi_{k,\tau_k} < \psi_{\ell,\tau_\ell} = 0$.

[46] This is the same indifference condition we found in case 1c as $\kappa \to \infty$, so aggregate information is continuous in this limit.

[47] In the borderline case with $p_0 = \bar{p}$, we get both $\Upsilon_k = 0$ and $\Upsilon_\ell = \infty$.

(54) leads to the contradiction that $\Upsilon_k = 0$.

Case 2b: $\rho \in (0,1)$. In this case we get an internal solution for $\Upsilon_k, \Upsilon_\ell$. We first further operationalize Lemma 11(b) by replacing the upper bound in (54) with an explicit expression for peripherals' pre-cutoff learning β_{ℓ,τ_ℓ} in terms of $\Upsilon_k, \Upsilon_\ell$, (57). To analyze (50) as the integrand $K^I a_t^I$ explodes and the integration domain $[\tau_k^I, \tau_\ell^I]$ vanishes, we rescale time $\mathfrak{a}_t^I := a_{t/I}^I$. The ODE (18) for core agents' experimentation intensity thus becomes

$$I\frac{\dot{\mathfrak{a}}_t^I}{1-\mathfrak{a}_t^I} = \begin{cases} L^I - 1 & t < I\tau_k^I \\ L^I - 1 - K^I\mathfrak{a}_t^I & t \in (I\tau_k^I, I\tau_\ell^I) \\ -K^I\mathfrak{a}_t^I & t > I\tau_\ell^I \end{cases} \tag{55}$$

Recalling $\rho, \Upsilon_k, \Upsilon_\ell$, as $I \to \infty$, the solution $\mathfrak{a}_t^I$ converges to the solution $\mathfrak{a}_t$ of

$$\frac{\dot{\mathfrak{a}}}{1-\mathfrak{a}} = \begin{cases} 1-\rho & t < \Upsilon_k \\ 1-\rho-\rho\mathfrak{a} & t \in (\Upsilon_k, \Upsilon_k + \Upsilon_\ell/(1-\rho)) \\ -\rho\mathfrak{a} & t > \Upsilon_k + \Upsilon_\ell/(1-\rho) \end{cases} \tag{56}$$

Peripherals' pre-cutoff learning (50) then converges to

$$\beta_{\ell,\tau_\ell} = \rho\left(\Upsilon_k + \int_{\Upsilon_k}^{\Upsilon_k+\Upsilon_\ell/(1-\rho)} \mathfrak{a}_t dt\right), \tag{57}$$

so we can rewrite Lemma 11(b) as

$$\Phi(\rho, \Upsilon_k, \Upsilon_\ell) := \rho\left(\Upsilon_k + \int_{\Upsilon_k}^{\Upsilon_k+\Upsilon_\ell/(1-\rho)} \mathfrak{a}_t dt\right) - \Upsilon_k = 0. \tag{58}$$

We can now characterize equilibrium learning.

Lemma 12. *For all $\rho \in [0,1]$, equations (53), (58) admit a unique solution $(\Upsilon_k, \Upsilon_\ell)$. This solution satisfies $0 < \Upsilon_k, \Upsilon_\ell < \infty$, and aggregate information $\Upsilon_k + \Upsilon_\ell$ falls in ρ.*

The proof of Lemma 12 relies on the following generalization of Leibniz's integral rule: For Lipschitz-continuous functions f, g and some cutoff $s > 0$, let x_t be the continuous solution of an ODE

$$\dot{x} = \begin{cases} f(x) & \text{for } t < s \\ g(x) & \text{for } t > s \end{cases}$$

with initial condition x_0. We write $x_t(s)$ to emphasize the importance of the cutoff, and

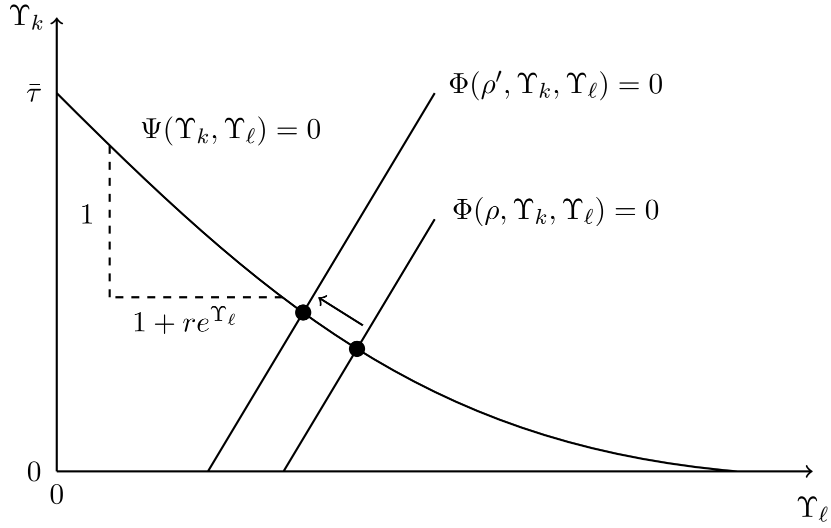

Figure 8: **Solutions of $\Phi(\rho, \Upsilon_k, \Upsilon_\ell) = 0$ and $\Psi(\Upsilon_k, \Upsilon_\ell) = 0$.**

assume $g(x_s(s)) \neq 0$.

Lemma 13. *For any $\Delta > 0$*

$$\frac{\partial}{\partial s} \int_s^{s+\Delta} x_t(s) dt = \frac{f(x_s(s))}{g(x_s(s))} \left(x_s(s+\Delta) - x_s(s)\right) \tag{59}$$

Proof of Lemma 12. Equation (58) together with $\Upsilon_k + \Upsilon_\ell > 0$ and the fact that the solution $\mathfrak{a}$ of (56) is bounded away from zero imply $\Upsilon_k > 0$, and in turn that $0 < \Upsilon_\ell < \infty$. Thus, asymptotic learning fails.

To solve (53), (58), we note that Φ clearly rises in ρ and Υ_ℓ. We show below that it falls in Υ_k. Hence zero-sets of Φ in $(\Upsilon_\ell, \Upsilon_k)$-space are increasing and shift left when ρ rises to ρ', as illustrated in Figure 8. Recalling from (52) that zero-sets of Ψ are decreasing with slope $-1/(1 + re^{-\Upsilon_\ell}) > -1$, equations (53), (58) admit a unique solution $(\Upsilon_k, \Upsilon_\ell)$. A rise in ρ shifts this solution left on the zero-set of Ψ, so $\Upsilon_k + \Upsilon_\ell$ falls.

In fact, the monotonicity of $\Upsilon_k + \Upsilon_\ell$ extends to the boundary points $\rho = 0, 1$: We recall that for $\rho = 0$ all learning is post-cutoff, $\Upsilon_k = 0, \Psi(0, \Upsilon_\ell) = 0$,[48] and anticipate that for $\rho = 1$ all learning is pre-cutoff, $\Upsilon_\ell = 0, \Psi(\Upsilon_k, 0) = 0$, thus attaining the extreme points on the zero set of $\Psi(\Upsilon_k, \Upsilon_\ell) = 0$ as illustrated in Figure 8.

To show that Φ falls in Υ_k, we write $\mathfrak{a}_* = \mathfrak{a}_{\Upsilon_k}$ and $\mathfrak{a}^* = \mathfrak{a}_{\Upsilon_k + \Upsilon_\ell/(1-\rho)}$, assume that

[48] This assumes $p_0 < \bar{p}$. For $p_0 \geq \bar{p}$, asymptotic information is infinite for $\rho = 0$, and hence trivially greater than the finite learning for $\rho > 0$.

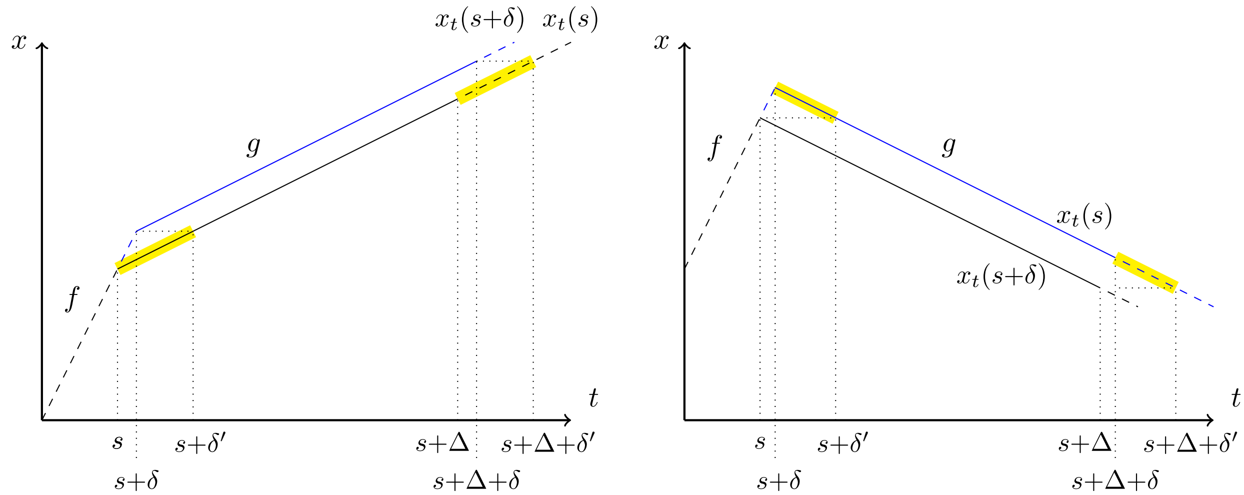

Figure 9: **Proof of Leibniz Rule.** In both figures the difference of between the integral of the upper solid line, $x_t(s+\delta)$ over $t \in [s+\delta, s+\delta+\Delta]$, and the lower solid line, $x_t(s)$ over $t \in [s, s+\Delta]$, equals the difference in the integrals of the shaded lines. E.g. in the left picture this is difference between $x_t(s)$ over $t \in [s+\Delta, s+\delta'+\Delta]$ and $x_t(s)$ over $t \in [s, s+\delta']$, which is the RHS of (61) after substituting $t = s + \tilde{\delta}$.

$1-\rho-\rho\mathfrak{a}_* \neq 0$, and then argue[49]

$$\frac{\partial \Phi}{\partial \Upsilon_k} = -(1-\rho) + \rho \frac{1-\rho}{1-\rho-\rho\mathfrak{a}_*}(\mathfrak{a}^* - \mathfrak{a}_*) = -(1-\rho)\frac{1-\rho-\rho\mathfrak{a}^*}{1-\rho-\rho\mathfrak{a}_*} < 0.$$

The first equality follows from Lemma 13 by substituting $s = \Upsilon_k$ and $\Delta = \Upsilon_\ell/(1-\rho)$ for the integral boundaries, $x_t = \mathfrak{a}_t$ for the trajectory, $f(\mathfrak{a}) = (1-\rho)(1-\mathfrak{a})$ for the law-of-motion before $s = \Upsilon_k$, and $g(\mathfrak{a}) = (1-\rho-\rho\mathfrak{a})(1-\mathfrak{a})$ after Υ_k. The middle equality is simple algebra, and the final inequality owes to the fact that $\dot{\mathfrak{a}}/(1-\mathfrak{a}) = 1-\rho-\rho\mathfrak{a}$ from (56) cannot switch signs on $[\Upsilon_k, \Upsilon_k + \Upsilon_\ell/(1-\rho)]$, so that $\frac{1-\rho-\rho\mathfrak{a}^*}{1-\rho-\rho\mathfrak{a}_*} > 0$. □

Proof of Lemma 13. The Leibniz rule evaluates the LHS of (59) "vertically", computing $\frac{\partial}{\partial s} x_t(s) = \lim_{\delta \to 0} \frac{1}{\delta}(x_t(s+\delta) - x_t(s))$ for fixed $t \in [s, s+\Delta]$. Since the ODE $\dot{x} = g(x)$ is autonomous, it is more economical to compare the trajectories $\{x_t(s+\delta)\}_t$ and $\{x_t(s)\}_t$ "horizontally", as illustrated in Figure 9.

Formally, assume first that $f(s)$ and $g(s)$ have the same sign, and for $\delta > 0$ small, let $\delta' > 0$ solve $x_{s+\delta'}(s) = x_{s+\delta}(s+\delta)$. At $s+\delta'$ the original trajectory "merges" with the shifted trajectory and since $\dot{x} = g(x)$ is autonomous we get $x_{s+\delta'+\hat{\delta}}(s) = x_{s+\delta+\hat{\delta}}(s+\delta)$, as illustrated

[49] Since $\mathfrak{a}_t = 1 - \exp(-(1-\rho)t)$ for $t < \Upsilon_k$, there exists at most one value of Υ_k with $1-\rho-\rho\mathfrak{a}_{\Upsilon_k} = 0$. Since Φ is continuous in Υ_k and decreasing in Υ_k everywhere else, it decreases everywhere.

in Figure 9(left). Thus

$$\int_{\delta}^{\delta+\Delta} x_{s+\tilde{\delta}}(s+\delta)d\tilde{\delta} = \int_{0}^{\Delta} x_{s+\delta+\hat{\delta}}(s+\delta)d\hat{\delta} = \int_{0}^{\Delta} x_{s+\delta'+\hat{\delta}}(s)d\hat{\delta} = \int_{\delta'}^{\delta'+\Delta} x_{s+\tilde{\delta}}(s)d\tilde{\delta} \tag{60}$$

using the change of variable $\tilde{\delta} = \delta + \hat{\delta}$ in the first equality, and $\tilde{\delta} = \delta' + \hat{\delta}$ in the last. Thus

$$\begin{aligned}\int_{s+\delta}^{s+\Delta+\delta} x_t(s+\delta)dt - \int_{s}^{s+\Delta} x_t(s)dt &= \int_{\delta}^{\delta+\Delta} x_{s+\tilde{\delta}}(s+\delta)d\tilde{\delta} - \int_{0}^{\Delta} x_{s+\tilde{\delta}}(s)d\tilde{\delta} \\ &= \int_{\delta'}^{\delta'+\Delta} x_{s+\tilde{\delta}}(s)d\tilde{\delta} - \int_{0}^{\Delta} x_{s+\tilde{\delta}}(s)d\tilde{\delta} = \int_{\Delta}^{\Delta+\delta'} x_{s+\tilde{\delta}}(s)d\tilde{\delta} - \int_{0}^{\delta'} x_{s+\tilde{\delta}}(s)d\tilde{\delta} \end{aligned} \tag{61}$$

where the first equality uses the change of variables $t = s + \tilde{\delta}$, the second uses (60), and the third cancels identical terms $\int_{\delta'}^{\Delta} x_{s+\tilde{\delta}}(s)d\tilde{\delta}$. In the limit

$$\frac{\partial}{\partial s}\int_{s}^{s+\Delta} x_t(s)dt = \lim_{\delta\to 0}\frac{\delta'}{\delta}\left(x_{s+\Delta}(s) - x_s(s)\right) = \frac{f(x_s(s))}{g(x_s(s))}\left(x_{s+\Delta}(s) - x_s(s)\right),$$

where we used that at first-order $\delta' g(x_s(s)) = \delta f(x_s(s))$.

If f and g have different signs, we let $\delta' > \delta$ solve $x_{s+\delta'}(s+\delta) = x_s(s)$, so $\delta f(s) + (\delta' - \delta)g(s) = 0$, as illustrated in Figure 9(right). Analogous arguments as above then show

$$\frac{\partial}{\partial s}\int_{s}^{s+\Delta} x_t(s)dt = \lim_{\delta\to 0}\frac{\delta'-\delta}{\delta}\left(x_s(s) - x_{s+\Delta}(s)\right) = \frac{f(x_s(s))}{g(x_s(s))}\left(x_{s+\Delta}(s) - x_s(s)\right). \quad \square$$

Case 2c: $\rho = 1$. While Lemmas 10 and 11 and most other substantive intermediate results remain true for $\rho = 1$, their proofs divide by $1-\rho$, and sometimes invoke that $L \to \infty$. Instead of re-proving everything, we provide a separate analysis, solely based on the function Ψ and its derivatives, (51)-(52), and the ODE (55). Specifically we will show that

$$\Psi(\Upsilon_k, \Upsilon_\ell) \le \psi_{k,\tau_k} \le 0 = \psi_{\ell,\tau_\ell} = \Psi(\Upsilon_k + \Upsilon_\ell, 0) \tag{62}$$

Together with (52), this implies $\Upsilon_\ell = 0$, so the inequalities in (62) must hold with equality. In particular $0 = \Psi(\Upsilon_k, 0) = P^{\emptyset}(\Upsilon_k)(x+y) - c$, so total information is as in the clique $\Upsilon_k + \Upsilon_\ell = \Upsilon_k = \bar{\tau}$.

We now show (62). The first inequality takes the limit of the strict inequality $\Psi(\Upsilon_k^I, \Upsilon_\ell^I) < \psi^I_{k,\tau_k^I}$, which reflects that core agents' observe post-cutoff information Υ_ℓ^I with a delay. The second inequality and the first equality reflect (the limits of) peripherals' indifference and core agents' weak shirking incentives at their respective cutoffs.

Only the last equality in (62), which states that peripherals' learning is entirely pre-cutoff, requires a novel argument and the assumption $\rho = 1$. Intuitively, information transmission by K^I core agents is infinitely faster than generation by L^I peripherals. Formally, we will show that peripherals' aggregate post-cutoff learning vanishes

$$\frac{K^I}{I}\int_{I\tau_\ell^I}^{\infty}\mathfrak{a}_t^I dt \to 0. \tag{63}$$

By (63), peripherals pre-cutoff learning $\beta^I_{\ell,\tau_\ell^I}$ converges to total information $\Upsilon_k+\Upsilon_\ell$, implying the last equality in (62).

To see (63) we first argue that $\mathfrak{a}_t^I \to 0$ for all t. By line one of (55), $\mathfrak{a}_t^I \le L^I t/I \le L^I\bar{\tau}/I \to 0$ for all $t < I\tau_k^I < \bar{\tau}$; at $t > I\tau_k^I$, lines two and three of (55) imply $\dot{\mathfrak{a}}_t^I < 0$ when $\mathfrak{a}_t^I \ge L^I/K^I \to (1-\rho)/\rho = 0$. All told, $\mathfrak{a}_t^I \to 0$ for all t. Turning to the aggregate in (63), line three of (55) states that $\mathfrak{a}_t^I$ decays exponentially at rate $(1-\mathfrak{a}_t^I)K^I/I$. Since this rate converges to 1, we have $\int_{I\tau_\ell^I}^{\infty}\mathfrak{a}_t^I dt - \mathfrak{a}^I_{I\tau_\ell^I} \to 0$. Together with $\mathfrak{a}^I_{I\tau_\ell^I} \to 0$, this implies (63).