Econometrica Supplementary Material

SUPPLEMENT TO "COMMUNICATION WITH UNKNOWN PERSPECTIVES"

(*Econometrica*, Vol. 84, No. 6, November 2016, 2029–2069)

BY RAJIV SETHI AND MUHAMET YILDIZ

IN THIS SUPPLEMENT, we consider some additional properties of long-run dynamics in Section A. Section B contains two variants of our model: a two-sided case in which the sets of observers and experts are disjoint, and a case with delayed observability of states. A model of shifting perspectives is explored in Section C.

A. FURTHER PROPERTIES OF LONG-RUN DYNAMICS

We write $b_{ij}^t = 1$ if the link ij is broken at t.

The following corollary of Proposition 1 establishes the frequency with which each $g \in G$ is realized in the long run, where $j_t = (j_{1t}, \dots, j_{nt})$ is the history-dependent network realized at time t.

COROLLARY 2: *If expertise levels are serially i.i.d., then, almost surely, the long-run frequency*

$$\phi_\infty(g|h) = \lim_{t\to\infty} \frac{\#\{s \le t | j_s(h) = g\}}{t} \quad (\forall g \in G)$$

exists, and

$$\phi_\infty(g|h) = P\Big(g(i) = \arg\max_{j\in J(i)} \pi_j \ \forall i \in N\Big).$$

When expertise levels are serially i.i.d., the realized networks are also i.i.d. in the long run, where the history-dependent long-run distribution is obtained by selecting the best-informed long-run expert for each i. This generates a testable prediction regarding the joint distribution of behavior in the long run: if both j and j' are elements of $J_h(i) \cap J_h(i')$, then i cannot link to j while i' links to j'. Furthermore, each pattern of linkage identified in Proposition 2 has an associated long-run distribution: long-run efficiency is characterized by an i.i.d. distribution on star networks, in which all players link to one player and that player links to another; the static network g is characterized by a point mass on g, and extreme opinion leadership is characterized by a point mass on a specific star network.

We now prove Corollary 2 and establish some additional results regarding long-run behavior. Let

$$D^\lambda = \big\{(\pi_1, \dots, \pi_n) \big| |\pi_i - \pi_j| \le \lambda\big\}$$

 DOI: 10.3982/ECTA13320

denote the set of expertise realizations such that each pair of expertise levels are within λ of each other. For any given J, let

$$p_{J,\lambda}(g) = \Pr\Big(g(i) = \arg\max_{j \in J(i)} \pi_j \; \forall i \in N \Big| \pi \notin D^{\lambda}\Big)$$

denote the conditional probability distribution on g obtained by restricting expertise realizations to lie outside the set D^{λ}. Finally, for any probability distribution p on G, let

$$B_{\varepsilon}(p) = \big\{q \big| \big|q(g) - p(g)\big| < \varepsilon \; \forall g \in G\big\}$$

denote the set of probability distributions q on G such that $q(g)$ and $p(g)$ are within ε of each other for all $g \in G$.

We say that $\phi_t(\cdot|h) \in B_{\varepsilon}(p)$ eventually if there exists $\bar{t}$ such that $\phi_t(\cdot|h) \in B_{\varepsilon}(p)$ for all $t > \bar{t}$. The following basic observations will also be useful in our analysis.

OBSERVATION 1: The following are true:

1. For every $\varepsilon > 0$, there exists $\lambda(\varepsilon) \in (0, \varepsilon)$ such that $\Pr(D^{\lambda(\varepsilon)}) < \varepsilon$.
2. For every $\lambda > 0$, there exists $\overline{v}_{\lambda} < \infty$ such that if $v_{ij}^t > \overline{v}_{\lambda}$ and $\pi_{jt} > \pi_{j't} + \lambda$, then $j_{it} \neq j'$.

The first of these observations follows from the fact that $\Pr(D^{\lambda})$ is continuous and approaches 0 as $\lambda \to 0$, and the second can be readily deduced using (5).

Next, we establish that, along every history, each link is eventually either broken or free. Define

$$\tilde{J}_h(i) = \Big\{j \Big| \lim_{t \to \infty} v_{ij}^t(h) > \overline{v}\Big\}$$

as the set of individuals j for which the link ij becomes free eventually. For any $J : N \to 2^N$ with $i \notin J(i)$, we also define

$$\tilde{H}^J = \big\{h_t | v_{ij}^t(h_t) > \overline{v} \text{ and } b_{ij'}(h_t) = 1 \; \big(\forall i \in N, \forall j \in J(i), \forall j' \notin J(i)\big)\big\}$$

as the set of histories in which all links ij with $j \in J(i)$ are free and all links ij' with $j' \notin J(i)$ are broken. We define $\tilde{H} = \bigcup_J \tilde{H}^J$ as the set of all histories at which all the links are resolved in the sense that they are either free or broken. Finally, we define the stopping time $\tilde{\tau}$ as the first time the process enters $\tilde{H}$, that is, $h_{\tilde{\tau}} \in \tilde{H}$ but $h_t \notin \tilde{H}$ for any $t < \tilde{\tau}$.

LEMMA 3: *The stopping time $\tilde{\tau}$ is finite, that is, for every h, there exists $\tilde{\tau}(h) < \infty$ such that $h_{\tilde{\tau}(h)} \in \tilde{H}$ but $h_t \notin \tilde{H}$ for all $t < \tilde{\tau}(h)$. Moreover, conditional on $h_{\tilde{\tau}}$,* *almost surely,*

$$J_h = \tilde{J}_h = \tilde{J}_{h_{\tilde{\tau}}},$$

where $\tilde{J}_{h_{\tilde{\tau}}}$ *is uniquely defined by* $h_{\tilde{\tau}} \in \tilde{H}^{\tilde{J}_{h_{\tilde{\tau}}}}$. *Finally,* $J_h = \tilde{J}_h$ *almost surely.*

PROOF: Consider any h. By definition, for every $i, j \in N$ with $j \in \tilde{J}_h(i)$, the link ij becomes free for the first time at some $\tau_{ij}(h)$. Moreover, by Lemma 1, for every $i, j \in N$ with $j \in J_h(i)$, we have $\lim_t v_{ij}^t(h_t) = \infty$. Hence, by Lemma 2, for every $j' \notin \tilde{J}$, the link ij is broken for the first time at some $\tau_{ij}(h)$.[10] Therefore, $h_{\tilde{\tau}(h)} \in \tilde{H}$ for the first time at $\tilde{\tau}(h) = \max_{i \in N, j \in J_h(i)} \tau_{ij}(h)$.

To prove the second part, observe that $\tilde{J}_h = \tilde{J}_{h_{\tilde{\tau}}}$ by definition. Moreover, $J_h \subseteq \tilde{J}_h$ because $\lim_t v_{ij}^t(h_t) = \infty$ whenever $j \in J_h(i)$. It therefore suffices to show that, conditional on $h_{\tilde{\tau}}$, each i links to each $j \in \tilde{J}_{h_{\tilde{\tau}}}(i)$ infinitely often almost surely. To establish this, take any i and j with $j \in \tilde{J}_{h_{\tilde{\tau}}}(i)$. Since $v_{ij}^{\tilde{\tau}}(h_{\tilde{\tau}}) > \overline{v}$, we have

$$\gamma\big(b, v_{ij}^{\tilde{\tau}}(h_{\tilde{\tau}})\big) < \gamma(b, \overline{v}) \le \gamma(a, v) \quad (\forall v),$$

where the first inequality is because γ is decreasing in $v_{ij}^{\tilde{\tau}}(h_{\tilde{\tau}})$ and the second is by definition of $\overline{v}$. Hence, by continuity of γ, there exists $\eta > 0$ such that

$$\gamma\big(b - \eta, v_{ij}^{\tilde{\tau}}(h_{\tilde{\tau}})\big) < \gamma(a + \eta, v) \quad (\forall v).$$

Since $v_{ij}^t(h_t) \ge v_{ij}^{\tilde{\tau}}(h_{\tilde{\tau}}) > \overline{v}$ for all continuations h_t of $h_{\tilde{\tau}}$, this further implies that

$$\gamma\big(b - \eta, v_{ij}^t(h_t)\big) < \gamma\big(a + \eta, v_{ik}^t(h_t)\big)$$

for every history h_t that follows $h_{\tilde{\tau}}$, for every k distinct from i and j, and for every t. Consequently, $l_{ij}^{t+1} = 1$ whenever $\pi_{jt} > b - \eta$ and $\pi_{kt} \le a + \eta$ for all other k. Thus,

$$\Pr\big(l_{ij}^{t+1} = 1\big) \ge F(a + \eta)^{n-2}\big(1 - F(b - \eta)\big) > 0$$

after any history that follows $h_{\tilde{\tau}}$ and any date $t \ge \tilde{\tau}$. Therefore, $l_{ij}^{t+1} = 1$ occurs infinitely often almost surely conditional on $h_{\tilde{\tau}}$. The last statement of the lemma immediately follows from the first two. *Q.E.D.*

Lemma 3 establishes that at some finite (history-dependent) time $\tilde{\tau}(h)$, all the links become either free or broken and remain so thereafter. That is when

[10] By Lemma 1 and definition of h, $\sup_t v_{ij'}^t(h) < \overline{v}$.

the set $\tilde{J}_h(i)$ of free links along infinite history h becomes known. Although the set $J_h(i)$ of long-run experts is contained in this set, some of the free links may not be activated after a while by chance. Lemma 3 establishes that such an event has zero probability, and all the free links are activated infinitely often. Eventually, all individuals learn the perspectives of their long-run experts to a high degree, and their behavior approaches their long-run behavior, with each individual linking to her most informed long-run expert.

Although the set of all long-run experts is known at time $\tilde{\tau}(h)$, it may take considerably longer for behavior to approach the long-run limit. Towards determining such time of convergence, for an arbitrary $\varepsilon > 0$, which will measure the level of approximation, and for any $J : N \to 2^N$ with $i \notin J(i)$, define the event

$$\hat{H}^{\lambda,J} = \{h_t | v^t_{ij}(h_t) > \overline{v}_\lambda \text{ and } b_{ij'}(h_t) = 1 \\ (\forall i \in N, \forall j \in J(i), \forall j' \notin J(i))\},$$

where $\overline{v}_\lambda$ is as in Observation 1. Define the event

$$H^\varepsilon = \bigcup_J \hat{H}^{\lambda(\varepsilon),J},$$

where $\lambda(\varepsilon)$ is as defined in Observation 1. When the process is in H^ε, we will have approximately the long-run behavior identified in Proposition 1. Define the stopping time $\hat{\tau}$ as the first time the process enters H^ε, that is, $h_{\hat{\tau}} \in H^\varepsilon$ but $h_t \notin H^\varepsilon$ for any $t < \hat{\tau}$. Define also $J_{h_{\hat{\tau}}}$ by $h_{\hat{\tau}} \in \hat{H}^{\lambda(\varepsilon),J_{h_{\hat{\tau}}}}$; this is well-defined because such $J_{h_{\hat{\tau}}}$ is unique. As discussed above, $\hat{\tau}$ may be infinite at some histories, but the total probability of such histories is zero by the last statement of Lemma 3. When $\hat{\tau}(h)$ is finite, we can take $\hat{\tau}(h)$ as $\tau(h)$ in Proposition 1. The next proposition summarizes our findings about long-run behavior.

PROPOSITION 6: *For every $\varepsilon \in (0, 1/n)$, there exists a set $\Pi \subset [a, b]^n$ with $\Pr(\pi_t \in \Pi) \geq 1 - \varepsilon$ such that, for all continuations h_t of all $h_{\hat{\tau}}$,*

1. $j_{it}(h_t, \pi_t) \in \{j \in J_{h_{\hat{\tau}}}(i) | \pi_{jt} \geq \pi_{j't} - \varepsilon \ \forall j' \in J_{h_{\hat{\tau}}}(i)\}$ *for all $i \in N$;*
2. $j_{it}(h_t, \pi_t) = \arg\max_{j \in J_{h_{\hat{\tau}}}(i)} \pi_{jt}$ *for all $i \in N$ whenever $\pi_t \in \Pi$;*
3. $|\Pr(j_t(h_t) = g) - p_{J_{h_{\hat{\tau}}}}(g)| \leq \varepsilon$ *for all $g \in G$;*
4. $J_h = J_{h_{\hat{\tau}}}$ *conditional on $h_{\hat{\tau}}$ almost surely.*

PROOF: Fix an arbitrary $\varepsilon > 0$, and set $\Pi = [a, b]^n \setminus D^{\lambda(\varepsilon)}$. Now, the first part of the proposition holds by definition. Indeed, for any continuation h_t of $h_{\hat{\tau}}$, $v^t_{ij}(h_t) \geq \overline{v}_{\lambda(\varepsilon)}$ whenever $j \in J_{h_{\hat{\tau}}}(i)$, and the link ij is broken whenever $j \notin J_{h_{\hat{\tau}}}(i)$. Hence, the statement follows from Observation 1 and from the fact that $\lambda(\varepsilon) < \varepsilon$. The second statement also immediately follows from the first. Now, since j_{it} differs from $\arg\max_{j \in J_{h_{\hat{\tau}}}(i)}$ only when $\pi_t \in D^{\lambda(\varepsilon)}$, we have

$$\left|\Pr(j_t(h_t) = g) - p_{J_{h_{\hat{\tau}}}}(g)\right| \leq \Pr(D^{\lambda(\varepsilon)}) < \varepsilon,$$

proving the third statement. To prove the fourth, for any $j \in J_{h_{\hat{\tau}}}(i)$, observe that $\Pr(j = \arg\max_{j' \in J_{h_{\hat{\tau}}}(i)} \pi_{j'}) = 1/|J_{h_{\hat{\tau}}}(i)| > 1/n$. Hence, by part 3, $\Pr(j_{it} = j | h_{\hat{\tau}}) > 1/n - \varepsilon > 0$ for all continuations. Therefore, by Kolmogorov's zero-one law, conditional on $h_{\hat{\tau}}$, $j_{it} = j$ infinitely often, that is, $j \in J_h(i)$, among any continuation h almost surely. *Q.E.D.*

Ignoring the zero probability event in which the set of long-run experts (determined by J_h) differs from the set of eventually free links (determined by $\tilde{J}_h$), Proposition 6 can be understood as follows. At some history-dependent time $\hat{\tau}$, all individuals learn the perspectives of all their long-run experts approximately. The first part states that they link to an approximately best-informed long-run expert thereafter. The second part states that they link to precisely the best-informed long-run expert with high probability. The third part states that, thereafter, the endogenous networks are approximately independently and identically distributed with $p_{J_{h_{\hat{\tau}}}}$, the distribution generated by selecting the most informed expert $j \in J_{h_{\hat{\tau}}}(i)$ for each i. Since $p_{J_{h_{\hat{\tau}}}}$ is history dependent, from an ex ante perspective the long-run exogenous networks are only exchangeable (i.i.d. with unknown distribution).

In the remainder of this section, we will prove Corollary 2, establishing the long-run frequency of endogenous networks. The following lemma is a key step.

LEMMA 4: *For any* $\lambda \in (0, 1)$, t_0, J, *and* $h_{t_0} \in \hat{H}^{\lambda,J}$ *and for any* $\varepsilon > \Pr(D^\lambda)$,

$$\Pr\big(\phi_t(\cdot|\cdot) \in B_\varepsilon(p_{J,\lambda}) \textit{ eventually} | h_{t_0}\big) = 1.$$

PROOF: For each $g \in G$ and each continuation history h of h_{t_0}, $\phi_t(g|h)$ can be decomposed as

$$\phi_t(g|h) = \phi_{t_0}(g|h_{t_0})\frac{t_0}{t} + \phi_{t,1}(g|h) + \phi_{t,2}(g|h),$$

where

$$\phi_{t,1}(g|h) = \frac{\#\{t_0 < s \le t | j_{is}(h) = g(i)\ \forall i \in N \text{ and } \pi_s \in D^\lambda\}}{t}$$

and

$$\phi_{t,2}(g|h) = \frac{\#\{t_0 < s \le t | j_{is}(h) = g(i)\ \forall i \in N \text{ and } \pi_s \notin D^\lambda\}}{t}$$
$$= \frac{\#\Big\{t_0 < s \le t | g(i) = \arg\max_{j \in J(i)} \pi_{js}\ \forall i \in N \text{ and } \pi_s \notin D^\lambda\Big\}}{t}.$$

Here, the last equality follows from the hypothesis in the lemma and the definition of $\overline{v}_\lambda$ in Observation 1. Hence, by the strong law of large numbers, as $t \to \infty$,

$$\begin{aligned}\phi_{t,2}(g|h) &\to \Pr\Big(g(i) = \arg\max_{j\in J(i)} \pi_{js}\ \forall i \in N \text{ and } \pi_s \notin D^\lambda\Big)\\ &= p_{J,\lambda}(g)\big(1 - \Pr\big(D^\lambda\big)\big),\end{aligned}$$

where the last equality is by definition. Thus, almost surely,

$$\begin{aligned}\limsup_t \phi_t(g|h) &= \limsup_t \phi_{t,1}(g|h) + p_{J,\lambda}(g)\big(1 - \Pr\big(D^\lambda\big)\big)\\ &\le p_{J,\lambda}(g) + \Pr\big(D^\lambda\big),\end{aligned}$$

where the inequality follows from the fact that $\limsup_t \phi_{t,1}(g|h) \le \Pr(D^\lambda)$, which in turn follows from the strong law of large numbers and the definition of $\phi_{t,1}$. Likewise, almost surely,

$$\begin{aligned}\liminf_t \phi_t(g|h) &= \liminf_t \phi_{t,1}(g|h) + p_{J,\lambda}(g)\big(1 - \Pr\big(D^\lambda\big)\big)\\ &\ge p_{J,\lambda}(g) - \Pr\big(D^\lambda\big),\end{aligned}$$

where the inequality follows from $\liminf_t \phi_{t,1}(g|h) \ge 0$ and $p_{J,\lambda}(g) \le 1$. Hence, for any $\varepsilon > \Pr(D^\lambda)$, for almost all continuations h of h_{t_0}, there exists $\overline{t}$ such that $\phi_t(g|h) \in (p_{J,\lambda}(g) - \varepsilon, p_{J,\lambda}(g) + \varepsilon)$ for all g. That is, $\phi_t(\cdot|h) \in B_\varepsilon(p_{J,\lambda})$ eventually, almost surely. *Q.E.D.*

PROOF OF COROLLARY 2: Ignore the zero probability event in which $\tilde{J}_h \neq J_h$ and $\hat{\tau}$ is infinite (see Lemma 3). Then, by the third part of Proposition 6, $J_h = J_{h_{\hat{\tau}}}$ almost surely, where $h_{\hat{\tau}}$ is the truncation of h to the time the process enters H^ε (along h). Define

$$\hat{H}^\varepsilon = \big\{h \in H | \phi_t(\cdot|h) \in B_{2\varepsilon}(p_{J_h}) \text{ eventually}\big\},$$

and observe that $\phi_t(\cdot|h) \in B_{2\varepsilon}(p_{J_h})$ whenever $\phi_t(\cdot|h) \in B_\varepsilon(p_{J_h,\lambda(\varepsilon)})$. But Lemma 4 states that, conditional on $h_{\hat{\tau}}$, $\phi_t(\cdot|h) \in B_\varepsilon(p_{J_h,\lambda(\varepsilon)})$ eventually with probability 1. That is, $\Pr(\hat{H}^\varepsilon|h_{\hat{\tau}}) = 1$ for each $h_{\hat{\tau}}$. Therefore,

$$\Pr\big(\hat{H}^\varepsilon\big) = 1.$$

Clearly, $\hat{H}^\varepsilon$ is increasing in ε, and as $\varepsilon \to 0$,

$$\hat{H}^\varepsilon \to \hat{H}^0 = \big\{h \in H | \phi_t(\cdot|h) \to p_{J_h}\big\}.$$

Therefore,

$$\Pr(\hat{H}^0) = \lim_{\varepsilon \to 0} \Pr(\hat{H}^\varepsilon) = 1. \qquad \textit{Q.E.D.}$$

B. VARIATIONS OF THE MODEL

The Two-Sided Case

Suppose that the set N of individuals is partitioned to two disjoint subsets: a set N_d of decision-makers, and a set N_e of potential experts. Only decision-makers make observational choices, and they can observe only potential experts. The domain and the range of graphs are modified accordingly; for example, $j_{it} \in N_e$, and it is defined only for $i \in N_d$. The definitions of the various patterns of long-run behavior are also adjusted accordingly. For example, opinion leadership is defined by $J_h(i) = \{j^*\}$ for all $i \in N_d$, and long-run efficiency is defined by $J_h(i) = N_e$ for all $i \in N_d$. In all other respects, the model is exactly as in the baseline case.

Our results concerning the behavior of a single individual clearly apply also to this variation. This includes our characterization of long-run behavior in Proposition 1, and our bound on the expected number of long-run experts in Proposition 3. The following result presents a crisper version of Proposition 2 for the two-sided model. In this version, within $(\tilde{v}, \overline{v} - \underline{\Delta})$, every graph emerges as a stable network with positive probability. Since the networks that involve segregation cannot arise outside of this region, this yields a sharp characterization.

PROPOSITION 7: *Under Assumption* 1, *for any* $v_0 \notin \{\tilde{v}, \overline{v}\}$, *the following are true*:

(a) *Long-run efficiency obtains with probability* 1 *if and only if* $v_0 > \overline{v}$.

(b) *Extreme opinion leadership emerges with positive probability if and only if* $v_0 < \overline{v}$, *and with probability* 1 *if and only if* $v_0 < \tilde{v}$.

(c) *For every* $v_0 \in (\tilde{v}, \overline{v} - \underline{\Delta})$, *every* $g : N_d \to N_e$ *emerges as a static network with positive probability.*

PROOF: The proofs of parts (a) and (b) are as in the one-sided model. The proof of part (c) is as follows. Fix any $v_0 \in (\tilde{v}, \overline{v} - \underline{\Delta})$, where $v_0 + \underline{\Delta} < \min\{\beta(v_0), \overline{v}\}$. By continuity of Δ and γ and by definition of β, there exists $\varepsilon \in (0, (b-a)/3)$ such that $v_0 + \Delta(a+\varepsilon, b-\varepsilon) < \min\{\beta(v_0), \overline{v}\}$,

$$\Delta(b-\varepsilon, b) > \Delta(a+\varepsilon, b-\varepsilon), \tag{25}$$

and

$$\gamma(b-\varepsilon, v_0) < \gamma\big(a+\varepsilon, v_0 + \Delta(a+\varepsilon, b-\varepsilon)\big). \tag{26}$$

Fix any such ε. Finally, fix any $g: N_d \to N_e$ and denote $g(N_d) = \{j_0, \dots, j_k\}$. Consider the following event Π: At any $t = 0, \dots, k$, the expertise levels of j_t and all i with $g(i) = j_t$ are greater than $b - \varepsilon$, and the expertise levels of all other individuals are less than $a + \varepsilon$. For $t = k+1, \dots, k+K$ with $v_0 + K\underline{\Delta} > \beta(v_0 + \Delta(a+\varepsilon, b-\varepsilon))$, all the expertise levels are in a neighborhood of the diagonal such that

$$\gamma\big(\pi_{jt}, v_0 + \Delta(b-\varepsilon, b)\big) < \gamma\big(\pi_{j't}, v_0 + \Delta(a+\varepsilon, b-\varepsilon)\big) \tag{27}$$

for all $j, j' \in N_e$. There is such an open nonempty neighborhood by (25). Now, at any $t = 0, \dots, k$, if $j_{it} = j_t$, then $v_{ij_{it}}^{t+1} > v_{ij_{it}}^t + \Delta(b-\varepsilon, b)$ when $g(i) = j_t$ and $v_{ij_{it}}^{t+1} < v_{ij_{it}}^t + \Delta(a+\varepsilon, b-\varepsilon)$ when $g(i) \neq j_t$. Hence, by (26), we have $j_{it} = j_t$ for all i with $g(i) = j_{k'}$ with $k' \geq t$, and $j_{it} \in \{g(i), j_t\}$ for all other i. Thus, $v_{ig(i)}^{k+1} > v_0 + \Delta(b-\varepsilon, b)$ and $v_{ij}^{k+1} < v_0 + \Delta(a+\varepsilon, b-\varepsilon)$ for all i and $j \neq g(i)$. Then, by (27), $j_{it} = g(i)$ for all i and all $t = k+1, \dots, k+K$. Therefore, all the links ij with $j \neq g(i)$ are broken at $k+K+1$, and $J_h(i) = \{g(i)\}$ for all i and all $h \in \Pi$. *Q.E.D.*

Using the ideas in the previous proofs, the following result delineates a subset of $(\tilde{v}, \overline{v})$ on which every nonempty correspondence $J: N_d \rightrightarrows N_e$ arises with positive probability. That is, one cannot say more than Proposition 1 about the behavior that arises with positive probability in the long run.

PROPOSITION 8: *Assume that there exists an integer m such that*

$$v_0 + m\underline{\Delta} < \overline{v}, \tag{28}$$

$$v_0 + m\Delta(b, b) > \overline{v}, \tag{29}$$

$$v_0 + \Delta(b, b) < \beta(v_0). \tag{30}$$

Then, for every nonempty $J: N_d \rightrightarrows N_e$, there exists a positive probability event on which $J_h = J$.

PROOF: Given the stated assumptions, there clearly exists $\varepsilon \in (0, (b-a)/3)$ such that

$$v_0 + m\Delta(a+\varepsilon, b-\varepsilon) < \overline{v}, \tag{31}$$

$$v_0 + m\Delta(b-2\varepsilon, b) > \overline{v}, \tag{32}$$

$$v_0 + \Delta(b-\varepsilon, b-\varepsilon) < \beta(v_0), \tag{33}$$

$$\gamma\big(a+\varepsilon, v_0 + \Delta(b-\varepsilon, b-\varepsilon)\big) > \gamma(b-\varepsilon, v_0), \tag{34}$$

where (33) and (34) follow from (30). Fix any nonempty correspondence $J: N_d \rightrightarrows N_e$, and define $J(N_d) = \bigcup_{i \in N_d} J(i)$. For the first $m|J(N_d)|$ periods, consider the periodic sequence of experts $j_t^* \in J(N_d)$ obtained by cycling through the members of $J(N_d)$. That is, j_1^* is the first member of $J(N_d)$, j_2^* is the second member of $J(N_d)$, …, $j_{|J(N_d)|}^*$ is the last member of $J(N_d)$, and $j_{k|J(N_d)|+t}^* = j_t^{*t}$. At any $t \leq m|J(N_d)|$, we have

$$\pi_{j,t} \in \begin{cases} [b-\varepsilon, b] & \text{if } j = j_t^*, \\ [b-2\varepsilon, b-\varepsilon) & \text{if } j \in J^{-1}(j^*) \equiv \{i \in N_e | J(i) = j^*\}, \\ [a, a+\varepsilon] & \text{otherwise.} \end{cases}$$

That is, j_t^* has the highest expertise, the decision-makers who would have j^* be a long-run expert according to J have the next highest levels of expertise, and all the other individuals have the lowest levels of expertise. Note that, by (33), in the first iteration of the cycle (the first $|J(N_d)|$ periods), we have $j_{it} = j_t^*$ for each $i \in N_e$. Since $\beta(v) - v$ is non-decreasing, this further implies that $j_{it} = j_t^*$ for each $i \in N_e$ at every $t \leq m|J(N_d)|$. Therefore, by the definition of m, at the end of period $m|J(N_d)|$, we have the link ij free (i.e., $v_{ij} > \overline{v}$) if and only if $j \in J(i)$. *Q.E.D.*

Observable States

Next, we consider the possibility that states are publicly observable with some delay. In particular, we assume that there exists $\tau \geq 0$ such that, for all t, θ_t becomes publicly observable at the end of period $t+\tau$. Note that $\tau = 0$ corresponds to observability of θ_t at the end of period t itself, as would be the case if one's own payoffs were immediately known. At the other extreme is the case where the state is never observed (as in our baseline model), which corresponds to the limit $\tau = \infty$.

With observable states, given any history at the beginning of date t, the precision of the belief of an individual i about the perspective of individual j is

$$v_{ij\tau}^t = v_{ij}^0 + \sum_{\{t' < t-\tau : j_{it'} = j\}} 1/\pi_{jt'} + \sum_{\{t-\tau \leq t' < t : j_{it'} = j\}} \Delta(\pi_{it'}, \pi_{jt'}). \quad (35)$$

For $t' < t - \tau$, individual i retrospectively updates her belief about the perspective of her target j at t' by using the true value of $\theta_{t'}$ instead of her private signal $x_{it'}$. This causes her belief about j's perspective to become more precise, rising by $1/\pi_{jt'}$ instead of $\Delta(\pi_{it'}, \pi_{jt'})$. Note that knowledge of the state does not imply knowledge of a target's perspective, since the target's signal remains unobserved.

This is the main effect of observability of past states: it retroactively improves the precision of beliefs about the perspectives of those targets who have been observed at earlier dates, without affecting the precision of beliefs about other individuals, along a given history. Such an improvement only enhances the attachment to previously observed individuals. This does not affect our results concerning one individual's behavior, such as the characterization of long-run behavior in Proposition 1 and the bound on the expected number of long-run experts in Proposition 3. Nor does it affect results concerning patterns of behavior that are symmetric on the observer side, such as long-run efficiency and opinion leadership in the first two parts of Proposition 2.[11]

Observability of states has a second effect, which relates to the asymmetry of observers. For $t' < t - \tau$, since an individual i already observes the true state $\theta_{t'}$, her signal $x_{it'}$ does not affect her beliefs at any fixed history, as seen in (35). Consequently, two individuals with identical observational histories have identical beliefs about the perspectives of all targets observed before $t - \tau$. This makes asymmetric linkage patterns, such as non-star-shaped static networks and information segregation, less likely to emerge. Nevertheless, when $\tau > 0$, individuals do use their private information in selecting targets until the state is observed. Therefore, under delayed observability, individuals' private signals impact their target choices, leading them to possibly different paths of observed targets. Indeed, our results about information segregation and static networks extend to the case of delayed observability for a sufficiently long delay τ.

Specifically, for a sufficiently long delay, every network emerges as the static network with positive probability. The reason for this is quite straightforward. Without observability, on a history under which g emerges as a static network, individuals become attached to their respective targets under g arbitrarily strongly over time. Hence, even if individuals start observing past states and learn more about other targets, the new information will not be sufficient to mend those broken links once enough time has elapsed. Moreover, for any partition of the population into sets of two or more individuals, there exists some $g \in G$ that maps each player i to a member of the same set in the partition. In this case, we have information segregation over the given partition.

To summarize, allowing for the observability of states with some delay does not alter the main message of this paper, and in some cases gives it greater force. The trade-off between being well-informed and being well-understood has interesting dynamic implications because those whom we observe become better understood by us over time. This effect is strengthened when a state is subsequently observed, since an even sharper signal of a target's perspective is obtained.

C. SHIFTING PERSPECTIVES

In our baseline model, we assumed that all perspectives were fixed: each individual assumes that θ_t is i.i.d. with a known distribution, and does not update

[11] To be precise, with observable states we have long-run efficiency whenever $v_0 > \overline{v}$, opinion leadership with positive probability when $v_0 < \overline{v}$, and opinion leadership with probability 1 when $v_0 < \tilde{v}$ (as in the first two parts of Proposition 2). However, the probability of opinion leadership may be 1 even when $v_0 > \tilde{v}$. Indeed, when $\tau = 0$, opinion leadership emerges with probability 1 whenever $v_0 < \tilde{v}'$, where $\tilde{v}' > \tilde{v}$ is defined by $\beta(\tilde{v}') - \tilde{v}' = 1/b$.

her beliefs about this distribution as she observes realizations of θ_t or signals about θ_t. We now consider the possibility that individuals recognize that they do not know the mean of θ_t and update their perspectives over time. We take

$$\theta_t = \mu + z_t, \tag{36}$$

where the random variables $(\mu, z_1, z_2, \dots)$ are stochastically independent and

$$\mu \sim_i N(\mu_{i0}, 1), \tag{37}$$

$$z_t \sim N(0, 1/\alpha_0). \tag{38}$$

Recall that $\sim_i$ indicates the belief of individual i, who believes that θ_t is i.i.d. with mean μ and variance 1, but does not know the mean μ; she believes—initially—that μ is normally distributed with mean μ_{i0} and precision α_0. We refer to the mean μ_{i0} as the initial perspective of i, and to the precision α_0 as the initial firmness of her perspective. We assume that θ_t is publicly observed at the end of the period t—as in the case of $\tau = 0$ in our discussion of observable states. This simplifies the analysis because individuals update their perspectives purely based on θ_t, rather than the signals and opinions they observe at t.

At the end of period t, the perspective of an individual i is

$$\mu_{it} \equiv E(\mu|\theta_1, \dots, \theta_t) = \frac{\alpha_0}{\alpha_0 + t}\mu_{i0} + \frac{t}{\alpha_0 + t}\frac{\theta_1 + \cdots + \theta_t}{t}, \tag{39}$$

and the firmness of her perspective (i.e., the precision of the belief about μ) is

$$\alpha_t \equiv \alpha_0 + t. \tag{40}$$

Note that the perspective μ_{it} is a convex combination of the initial perspective μ_{i0} and the empirical average $\overline{\theta}_t = (\theta_1 + \cdots + \theta_t)/t$ of the realized states with deterministic weights α_0/α_t and t/α_t, respectively.[12] As time progresses, the weight α_0/α_t of the initial perspective decreases and eventually vanishes, while the weight t/α_t of the empirical average approaches 1.

Individuals other than i have two sources of information about the revised perspective μ_{it}: (i) the past opinions of i, which are endogenously and privately observed and are the only sources of information about the initial perspective μ_{i0}, and (ii) the realization of past states, which are exogenously and publicly observed and provide information about the data that individual i uses to update her perspective.

In earlier periods, their main information comes endogenously from the first source, as in our baseline model. However, in the long run, the accumulated

[12]Note also that the perspective μ_{it} is a random variable, as it depends on an empirical average, while its firmness α_t is deterministic and increasing in t.

data coming from the public source dominates the privately obtained information, as the perspective approaches the empirical average of the realized states. They eventually learn the perspective of each individual so precisely that they choose their targets based purely on expertise levels, as in the case of known perspectives. Hence efficiency is the only possible outcome in the long run. Nevertheless, the speed of convergence is highly dependent on the initial firmness α_0 of each perspective. The long-run behavior can be postponed indefinitely by considering firmer and firmer initial perspectives. Moreover, under such firm perspectives, the belief dynamics will also be similar to those in our baseline model, which corresponds to infinite initial firmness. Hence, the behavior will be similar to the long-run behavior in the baseline model in those arbitrarily long stretches of time before the perspectives are learned—as we establish below.

Effect of Learning on Choosing Targets

We next describe how individuals choose their targets. We show that, in comparison with the baseline model, there is a stronger motive to listen to better-understood targets vis-à-vis better-informed ones, and this motive decreases over time and approaches the baseline model in the limit. That is, learning strengthens path dependence early on.

At the beginning of period t, the belief of any individual j about the state is

$$\theta_t \sim_j N(\mu_{j(t-1)}, 1 + 1/\alpha_{t-1}).$$

This is as in our baseline model, except that the individual faces additional uncertainty about the underlying distribution, so the variance is $1 + 1/\alpha_{t-1}$ rather than 1. Her opinion is accordingly

$$y_{jt} = \frac{1}{1+\hat{\pi}_{jt}}\mu_j + \frac{\hat{\pi}_{jt}}{1+\hat{\pi}_{jt}}x_{jt},$$

where

$$\hat{\pi}_{jt} = \pi_{jt}/(1 + 1/\alpha_{t-1}). \tag{41}$$

That is, opinions are formed as in our baseline model, but with individuals having effectively lower expertise, reflecting their uncertainty about the underlying process. The effective expertise level $\hat{\pi}_{jt}$ approaches the nominal expertise level π_{jt} of our baseline model as $t \to \infty$. This modification can be incorporated into our earlier analysis by modifying the distribution of expertise levels at each period, taking the bounds of expertise to be

$$a_t = a/(1 + 1/\alpha_{t-1}) \quad \text{and} \quad b_t = b/(1 + 1/\alpha_{t-1}), \tag{42}$$

which converge to the original bounds a and b, respectively, as $t \to \infty$.

Writing v_{ij}^t for the precision of the belief of i about the perspective $\mu_{j(t-1)}$ of player j at the beginning period t, the opinion y_{jt} provides a noisy signal

$$\frac{1+\hat{\pi}_{jt}}{\hat{\pi}_{jt}} y_{jt} = \theta_t + \varepsilon_{jt} + \frac{1}{\hat{\pi}_{jt}} \mu_{j(t-1)},$$

as in the baseline model. Once again, the variance of the additive noise in the signal observed by i is

$$\gamma(\hat{\pi}_{jt}, v_{ij}^t) \equiv \frac{1}{\hat{\pi}_{jt}} + \frac{1}{\hat{\pi}_{jt}^2} \frac{1}{v_{ij}^t}.$$

This leads to the same behavior as in the baseline model, with effective expertise replacing nominal expertise:

$$j_{it} = \min\Big\{\arg\min_{j \neq i} \gamma(\hat{\pi}_{jt}, v_{ij}^t)\Big\}. \tag{43}$$

Here, individual i simply discounts the expertise levels of her potential targets, making expertise less valuable vis-à-vis familiarity, tilting the scale towards better-understood targets. That is, in the short run, learning actually increases the attachment to previously observed targets, leading to stronger path-dependence. Towards stating this formally, we define the marginal rate of substitution of expertise level π_{jt} for the precision v_{ij}^t of variance at t as

$$MRS_{\pi,v}^t \equiv -\frac{\partial\gamma(\hat{\pi}_{jt}, v_{ij}^t)/\partial\pi_{jt}}{\partial\gamma(\hat{\pi}_{jt}, v_{ij}^t)/\partial v_{ij}^t} = \frac{1}{1+\alpha_{t-1}} + 2v_{ij}^t/\partial\pi_{jt}.$$

In the baseline model, the marginal rate of substitution is

$$\overline{MRS}_{\pi,v} \equiv -\frac{\partial\gamma(\pi_{jt}, v_{ij}^t)/\partial\pi_{jt}}{\partial\gamma(\pi_{jt}, v_{ij}^t)/\partial v_{ij}^t} = 1 + 2v_{ij}^t/\partial\pi_{jt}.$$

The following proposition immediately follows from the above expressions.

PROPOSITION 9: *The marginal rate of substitution of π_{jt} for the precision v_{ij}^t of variance is higher in the model with learning*:

$$MRS_{\pi,v}^t > \overline{MRS}_{\pi,v}.$$

Moreover, $MRS_{\pi,v}^t$ is decreasing in t and converges to $\overline{MRS}_{\pi,v}$ as $t \to \infty$. In particular, for any i, j, and j' with fixed $v_{ij}^t > v_{ij'}^t$, if i prefers j to j' at t in the baseline model, she also prefers j to j' at t in the model with learning.

That is, learning increases attachment to familiar targets at the expense of more informed ones when we fix beliefs about the other players' perspectives. However, the beliefs about the other players' perspectives are different under learning—as we show next.

Belief Dynamics

The updating of beliefs about perspectives is somewhat more interesting. At any history h at the beginning of a date t, write $v_{ij}^t = 1/\operatorname{Var}(\mu_{j(t-1)}|h)$ and $v_{ij0}^t(h) = 1/\operatorname{Var}(\mu_{j0}|h)$ for the precisions of the beliefs of i about the current and the initial perspective of j, respectively. By (39), the variance of the current perspective $\mu_{j(t-1)}$ is

$$\operatorname{Var}(\mu_{j(t-1)}|h) = \frac{1}{v_{ij0}^t(h)}\left(\frac{\alpha_0}{\alpha_t}\right)^2 + \frac{1}{t}\left(\frac{t}{\alpha_t}\right)^2. \tag{44}$$

Here, the first term reflects the uncertainty about the initial perspective and depends on previous observations. The second term reflects the flow of public information, depending only on time and the firmness of beliefs. Then, the precision of beliefs about the current perspective is

$$v_{ij}^t = 1/\operatorname{Var}(\mu_{j(t-1)}|h) = \frac{\alpha_t^2 v_{ij0}^t(h)}{\alpha_0^2 + t v_{ij0}^t(h)}. \tag{45}$$

Since $\alpha_t = \alpha_0 + t$, we observe that v_{ij}^t is an increasing function of both v_{ij0}^t (reflecting the information gathered by observing the opinions directly) and t (reflecting the flow of public information).

To determine v_{ij}^t, we next determine v_{ij0}^t. As in the baseline model, observation of y_{jt} by i provides the following signal about $\mu_{j(t-1)}$:

$$(1 + \hat{\pi}_{jt})y_{jt} - \hat{\pi}_{jt}\theta_t = \mu_{j(t-1)} + \hat{\pi}_{jt}\varepsilon_{jt}.$$

Together with (39), this leads to the following signal about μ_{j0}:

$$\frac{\alpha_{t-1}}{\alpha_0}(1 + \hat{\pi}_{jt})y_{jt} - \left[\frac{\theta_1 + \cdots + \theta_{t-1}}{\alpha_0} + \frac{\alpha_{t-1}}{\alpha_0}\hat{\pi}_{jt}\theta_t\right] = \mu_{j0} + \frac{\alpha_{t-1}}{\alpha_0}\hat{\pi}_{jt}\varepsilon_{jt}.$$

The precision of this signal (i.e., the inverse of the variance of the additive noise term $\frac{a_{t-1}}{\alpha_0}\hat{\pi}_{jt}\varepsilon_{jt}$) is

$$\Delta_t(\pi_{jt}) = \frac{\alpha_0^2(1 + \alpha_{t-1})^2}{\alpha_{t-1}^4}\frac{1}{\pi_{jt}}.$$

As in the baseline model, this leads to the following closed-form solution:

$$v_{ij0}^t = v_0 + \sum_{s=1}^{t-1} \Delta_s(\pi_{js}) l_{ij}^s, \tag{46}$$

where l_{ij}^s is 1 if i links to j at s and 0 otherwise. By substituting (46) into (45), we also obtain a formula for v_{ij}^t.

Note that one applies to each variance $1/\pi_{js}$ a weight that is decreasing in s, where the weight is approximately $(\alpha_0/\alpha_{s-1})^2$. That is, earlier observations add more precision to the belief about μ_{j0}—because those opinions reflect the initial perspective more strongly. As time progresses, the opinions become less valuable sources of information about the initial perspective, and their impact on v_{ij0}^{t+1} eventually becomes negligible. Note also that the precisions v_{ij0}^t are uniformly bounded from above.

Long-Run Behavior and the Robustness of the Baseline Model

Since the precisions v_{ij0}^t are uniformly bounded, the long-run behavior is driven by the flow of public information. Indeed, by (45), when $t \gg \alpha_0$, v_{ij}^t is approximately t, regardless of history. Hence, for any fixed α_0, as $t \to \infty$, v_{ij}^t also goes to ∞, leading to long-run efficiency.

PROPOSITION 10: *In the model with learning, for any fixed α_0, we have long-run efficiency.*

PROOF: Since $v_{ij0}^t \geq v_0$ and v_{ij}^t is an increasing function of v_{ij0}^t, we can conclude from (45) that

$$v_{ij}^t(h) \geq \frac{\alpha_t^2 v_0}{\alpha_0^2 + t v_0} = \frac{(\alpha_0 + t)^2 v_0}{\alpha_0^2 + t v_0} \quad (\forall h \in H).$$

Hence, for every history $h \in H$,

$$\lim_{t\to\infty} v_{ij}^t(h) \geq \lim_{t\to\infty} \frac{(\alpha_0 + t)^2 v_0}{\alpha_0^2 + t v_0} = \infty.$$

Q.E.D.

Despite this, the long-run outcome can be postponed indefinitely by choosing firmer initial beliefs, and the medium-run behavior is similar to the long-run behavior of our baseline model. This can be deduced from (45) and (46) as follows. In (45), for any fixed (t, v_{ij0}^t), $\lim_{\alpha_0\to\infty} v_{ij}^t = v_{ij0}^t$. At the same time, in (46),

$$\lim_{\alpha_0\to\infty} v_{ij0}^t = v_0 + \sum_{s=1}^{t-1} l_{ij}^s/\pi_{js} \equiv \overline{v}_{ij}^t,$$

where $\overline{v}_{ij}^t$ is the precision of the belief of i about the perspective of j in the baseline model with observable states. Hence,

$$\lim_{\alpha_0 \to \infty} v_{ij}^t = v_0 + \sum_{s=1}^{t-1} l_{ij}^s / \pi_{js} \equiv \overline{v}_{ij}^t. \tag{47}$$

That is, the beliefs as functions of past behavior remain close to those in the baseline model when the initial conviction in perspectives is sufficiently high. Since past behavior is also a function of past beliefs, this further implies that both behavior and beliefs remain close to their counterparts in the baseline model. This is formalized in the next result. The first part states that, with high probability, individuals choose their target between dates $\overline{t}$ and $\overline{t}+l$ according to the long-run behavior without learning. The second part states that the possible patterns of behavior within that (arbitrarily long) time interval coincide with the patterns possible under the long-run behavior of the baseline model.

PROPOSITION 11: *For every $\varepsilon > 0$ and any positive integer l, there exist a date $\overline{t} < \infty$ and $\overline{\alpha} < \infty$ such that*

$$\Pr\Big(j_{it}(h) \in \arg\max_{j \in \overline{J}_h(i)} \pi_{jt} \ \forall t \in \{\overline{t}, \overline{t}+1, \dots, \overline{t}+l\} \Big) > 1 - \varepsilon \quad (\forall \alpha_0 > \overline{\alpha}),$$

where $\overline{J}_h(i)$ is the set of long-run experts in the baseline model with observable states. Moreover, for any mapping $J : N \to 2^N \setminus \{\emptyset\}$, we have $\Pr(\overline{J}_h = J) > 0$ if and only if

$$\Pr\Big(j_{it}(h) \in \arg\max_{j \in \overline{J}_h(i)} \pi_{jt} \ \forall t \in \{\overline{t}, \overline{t}+1, \dots, \overline{t}+l\} \big| h_{\overline{t}} \in H^J \Big) > 1 - \varepsilon$$

$$(\forall \alpha_0 > \overline{\alpha})$$

for some positive probability event H^J.

PROOF: Fix some positive ε and l as in the proposition. There then exists $\varepsilon' > 0$ such that $\Pr(\Pi^{\varepsilon'}) > (1 - \varepsilon/4)^{1/l}$ where $\Pi^{\varepsilon'} = \{\pi || \pi_i - \pi_j| > \varepsilon'\}$. There also exists finite $\overline{t}$ such that $\Pr(H') > 1 - \varepsilon/4$ for $H' = \{h \in H | \tau(h) < \overline{t}\}$ where the $\tau(h)$ is defined for $\varepsilon'/2$ in Proposition 1. Note that there exists $\lambda > 0$ such that whenever $(\pi_{1t}, \dots, \pi_{nt}) \in \Pi^{\varepsilon'}$ and $t \geq \overline{t} > \tau(h)$, we have $\gamma(\pi_{j_{it}^*(h)t}, \overline{v}_{ij_{it}^*(h)}^t(h)) < \gamma(\pi_{jt}, \overline{v}_{ij}^t(h)) - \lambda$ for all $j \in N \setminus \{i, j_{it}^*(h)\}$ where $j_{it}^*(h) = \arg\max_{j \in \overline{J}_h(i)} \pi_{jt}$ and $\overline{v}_{ij}^t(h)$ is the precision of belief of i about the perspective of j in the baseline model under observable states. One can further show that there exists $\lambda' \in (0, \lambda)$ such that the probability of the event

$$H'' = \big\{h \in H \mid \gamma\big(\pi_{j^*_{it}(h)t}, \overline{v}^t_{ij^*_{it}(h)}(h)\big) < \gamma\big(\pi_{jt}, \overline{v}^t_{ij}(h)\big) - \lambda' \quad \forall t \leq \overline{t}, j \in N \setminus \big\{i, j^*_{it}(h)\big\}\big\}$$

also exceeds $1 - \varepsilon/4$. Consider the set $\hat{H}$ of histories in the intersection of the events H', H'', and that all the realizations of expertise levels between $\overline{t}$ and $t + l$ are in $\Pi^{\varepsilon'}$. Clearly, $\Pr(\hat{H}) > 1 - \varepsilon$, as the probabilities of the excluded events add up to $3\varepsilon/4$. Note, however, that, since $v^t_{ij}(h) \geq v_0$ throughout, there then exists $\lambda'' > 0$ such that, for all

$$\big\| v^t_i(h) - \overline{v}^t_i(h) \big\| < \lambda'' \quad \Rightarrow \quad \gamma\big(\pi_{j^*_{it}(h)t}, v^t_{ij^*_{it}(h)}(h)\big) < \gamma\big(\pi_{jt}, v^t_{ij}(h)\big) \quad \forall h \in \hat{H}, t \leq \overline{t} + l, j \in N \setminus \big\{i, j^*_{it}(h)\big\}.$$

But, since the limit in (47) is uniform over all histories with $t \leq \overline{t} + l$, there also exists $\overline{\alpha}$ such that $\|v^t_i(h) - \overline{v}^t_i(h)\| < \lambda''$ for all histories with $t \leq \overline{t} + l$ whenever $\alpha_0 > \overline{\alpha}$.

The second part of the proposition can be obtained from the first part using Proposition 6. *Q.E.D.*

Dept. of Economics, Barnard College, Columbia University, 3009 Broadway, New York, NY 10027, U.S.A.; rs328@columbia.edu

and

Dept. of Economics, MIT, 50 Memorial Drive, Cambridge, MA 02142, U.S.A.; myildiz@mit.edu.

Co-editor Joel Sobel handled this manuscript.

Manuscript received March, 2015; final revision received May, 2016.