Appendix: Only for online publication

Contents

A **Exit problem: one population models** **35**

B **Exit problem: two-population models** **43**

C **Stochastic stability: the maximin criterion** **50**

D **Stochastic stable states for Nash demand games** **53**

A. Exit problem: one population models

Proof of Lemma 4.1. (i) Since $c^{(n)}(x, x^{i,k}) = \pi(\bar{m}, x) - \pi(k, x)$, we obtain

$$\begin{aligned}
I^{(n)}(\gamma_2) - I^{(n)}(\gamma_1) &= [\pi(\bar{m}, x) - \pi(k, x) + \pi(\bar{m}, x^{i,k}) - \pi(l, x^{i,k})] \\
&\quad - [\pi(\bar{m}, x) - \pi(k, x) + \pi(\bar{m}, x^{\bar{m},k}) - \pi(l, x^{\bar{m},k})] \\
&= \frac{1}{n}\left([-A_{\bar{m}i} + A_{\bar{m}k} + A_{li} - A_{lk}] - [-A_{\bar{m}\bar{m}} + A_{\bar{m}k} + A_{l\bar{m}} - A_{lk}]\right) \\
&= \frac{1}{n}\left(A_{\bar{m}\bar{m}} - A_{l\bar{m}} - A_{\bar{m}i} + A_{li}\right) > 0
\end{aligned}$$

from the **MBP**.
(ii) We find that

$$\begin{aligned}
&[I^{(n)}(\zeta_2) - I^{(n)}(\zeta_1)] + [I^{(n)}(\zeta_2) - I^{(n)}(\zeta_3)] \\
=&[\pi(\bar{m}, x) - \pi(i, x) + \pi(\bar{m}, x^{\bar{m},i}) - \pi(j, x^{\bar{m},i}) + \pi(\bar{m}, x^{(\bar{m},i)(\bar{m},j)}) - \pi(i, x^{(\bar{m},i)(\bar{m},j)})] \\
-&[\pi(\bar{m}, x) - \pi(j, x) + \pi(\bar{m}, x^{\bar{m},j}) - \pi(i, x^{\bar{m},j}) + \pi(\bar{m}, x^{(\bar{m},j)(\bar{m},i)}) - \pi(i, x^{(\bar{m},j)(\bar{m},i)}))] \\
+&[\pi(\bar{m}, x) - \pi(i, x) + \pi(\bar{m}, x^{\bar{m},i}) - \pi(j, x^{\bar{m},i}) + \pi(\bar{m}, x^{(\bar{m},i)(\bar{m},j)}) - \pi(i, x^{(\bar{m},i)(\bar{m},j)})] \\
-&[\pi(\bar{m}, x) - \pi(i, x) + \pi(\bar{m}, x^{\bar{m},i}) - \pi(i, x^{\bar{m},i}) + \pi(\bar{m}, x^{(\bar{m},i)(\bar{m},i)}) - \pi(j, x^{(\bar{m},i)(\bar{m},i)})] \\
=&[A_{\bar{m}i} - A_{\bar{m}j} + A_{j\bar{m}} - A_{ji} - A_{i\bar{m}} + A_{ij}] + [A_{i\bar{m}} - A_{ij} + A_{\bar{m}j} - A_{\bar{m}i} - A_{j\bar{m}} + A_{ji}] \\
=&0
\end{aligned}$$

From this we obtain the desired results. □

Proof of Proposition 4.1. **Part (i).** In the proof, we suppress the superscript (n). Let $\gamma = (x_1, x_2, \cdots, x_T)$ be a path in $\mathcal{G}_{\bar{m}} \setminus \mathcal{J}_{\bar{m}}$. We recursively construct a new path $\tilde{\gamma} \in \mathcal{J}_{\bar{m}}$ with a cost lower than or equal to the cost of γ.

For this, let t be the greatest number such that $x_{t+1} = (x_t)^{i,l}$ with $i \neq \bar{m}, l$. We distinguish several cases. If $t = T-1$, we consider a new path $\tilde{\gamma}$ obtained by modifying the last transition as follows:

$$\tilde{\gamma} := (x_1, x_2, \cdots, x_{T-1}, (x_{T-1})^{\bar{m},l}).$$

Then, we have $I(\tilde{\gamma}) = I(\gamma)$, and show that the path still exits $D(e_{\bar{m}})$. To prove this, we only need to show that if $z \notin D(e_{\bar{m}})$ then $z^{\bar{m},i} \notin D(e_{\bar{m}})$, because this implies that if $(x_{T-1})^{i,l} \notin D(e_{\bar{m}})$, then $(x_{T-1})^{\bar{m},l} \notin D(e_{\bar{m}})$. Now, suppose that $z \notin D(e_{\bar{m}})$ and that there exists k such that $\pi(\bar{m}, z) < \pi(k, z)$. Then, we have

$$[\pi(k, z^{\bar{m},i}) - \pi(\bar{m}, z^{\bar{m},i})] - [\pi(k, z) - \pi(\bar{m}, z)] = \frac{1}{n}\left(A_{ki} - A_{k\bar{m}} - A_{\bar{m},i} + A_{\bar{m},\bar{m}}\right) \geq 0$$

by Condition **A**. Thus, we have $[\pi(k, z^{\bar{m},i}) - \pi(\bar{m}, z^{\bar{m},i})] \geq [\pi(k, z) - \pi(\bar{m}, z)] > 0$ and so $z^{\bar{m},i} \notin D(e_{\bar{m}})$.

Now, suppose that $t < T-1$. Then we have $x_{t+1} = (x_t)^{i,l}$ and $x_{t+2} = (x_t)^{(i,l)(\bar{m},k)}$ for $k \neq \bar{m}$. Note that $k \neq \bar{m}$ and $l \neq i$. Now we need to distinguish four cases.
Case 1: If $k = i, l = \bar{m}$, then $x_{t+1} = (x_t)^{i,\bar{m}}, x_{t+2} = x_t$. Thus, we consider $\tilde{\gamma} = (x_1, \cdots, x_t, x_{t+2}, \cdots, x_T)$; clearly, $I(\tilde{\gamma}) \leq I(\gamma)$, since $c(x_t, x_{t+1}) = 0, c(x_{t+1}, x_{t+2}) \geq 0,$ and $c(x_t, x_{t+2}) = 0$.

Case 2: If $k = i, l \neq \bar{m}$ then $x_{t+2} = (x_t)^{(i,l)(\bar{m},k)} = (x_t)^{\bar{m},l}$. Again, we consider the path $\tilde{\gamma} = (x_1, \cdots, x_t, x_{t+2}, \cdots, x_T)$ and find that $I(\tilde{\gamma}) \leq I(\gamma)$ because we have $c(x_t, x_{t+1}) = c(x_t, x_{t+2}) = \pi(m, x_t) - \pi(l, x_t)$ and $c(x_{t+1}, x_{t+2}) \geq 0$.
Case 3: If $k \neq i, l = \bar{m}$, then $x_{t+2} = x_t^{(i,\bar{m})(\bar{m},k)} = (x_t)^{i,k}$. Again, let $\tilde{\gamma} = (x_1, \cdots, x_t, x_{t+2}, \cdots, x_T)$. Then we have $c(x_t, x_{t+1}) = 0$ and

$$\begin{aligned} c(x_{t+1}, x_{t+2}) - c(x_t, x_{t+2}) &= c(x_t^{i,l}, x_t^{(i,l)(\bar{m},k)}) - c(x_t, x_t^{(i,k)}) \\ &= \pi(\bar{m}, x_t^{i,\bar{m}}) - \pi(k, x_t^{i,\bar{m}}) - [\pi(\bar{m}, x_t) - \pi(k, x_t)] \\ &= \frac{1}{n}(A_{\bar{m}\bar{m}} - A_{k\bar{m}} - [A_{\bar{m}i} - A_{ki}]) \geq 0 \end{aligned}$$

from the **MBP**, implying that $I(\tilde{\gamma}) \leq I(\gamma)$.
Case 4: If $k \neq i, \bar{m}$ and $l \neq i, \bar{m}$, then we can apply Lemma 4.1. We modify the path by considering the alternative transitions, $\tilde{x}_{t+1} = (x_t)^{\bar{m},l}$ and $\tilde{x}_{t+2} = (x_t)^{(\bar{m},l)(i,k)}$. If $(x_t)^{\bar{m},l} \notin D(e_{\bar{m}})$, then we define

$$\tilde{\gamma} := (x_1, x_2, \cdots, x_t, (x_t)^{\bar{m},l})$$

and because $c(x_t, (x_t)^{\bar{m},l}) = c(x_t, (x_t)^{i,l})$ and $c(x_{t+1}, x_{t+2}) \geq 0$, we obtain $I(\tilde{\gamma}) \leq I(\gamma)$. If $(x_t)^{\bar{m},l} \in D(e_{\bar{m}})$, then we define

$$\tilde{\gamma} := (x_1, x_2, \cdots, x_t, (x_t)^{\bar{m},l}, (x_t)^{(\bar{m},l)(i,k)}, \cdots, x_T).$$

to find that $I(\tilde{\gamma}) \leq I(\gamma)$ from Lemma 4.1. Proceeding inductively we construct a path $\tilde{\gamma} \in \mathcal{J}_{\bar{m}}$ with a cost lower than or equal to the cost of γ.

Part (ii). We denote by $c(a, a^{i,j,\rho})$ be the cost of a path from a to $a^{i,j,\rho}$ in which agents switch from i to j, ρ-times consecutively and let $\pi(k, x - y) := \pi(k, x) - \pi(k, y)$ and $\gamma_{a \to b}$ be a path from a to b. We first show the following lemma.

Lemma A.1. *We have the following results.*

$$\begin{aligned} &(i) \quad c(a, a^{\bar{m},k,\rho}) - c(b, b^{\bar{m},k,\rho}) = \rho[(\pi(\bar{m}, a) - \pi(k, a)) - (\pi(\bar{m}, b) - \pi(k, b))] \\ &(ii) \quad \eta[c(a, a^{\bar{m},k,\rho}) - c(b, b^{\bar{m},k,\rho})] + \rho[c(b^{k,\bar{m},\eta}, b) - c(a^{k,\bar{m},\eta}, a)] = 0 \\ &(iii) \quad \eta[I(\gamma_{a^{\bar{m},k,\rho} \to b^{\bar{m},k,\rho}}) - I(\gamma_{a \to b})] + \rho[I(\gamma_{a^{k,\bar{m},\eta} \to b^{k,\bar{m},\eta}}) - I(\gamma_{a \to b})] = 0 \end{aligned}$$

where $\gamma_{a^{k,\bar{m},\eta} \to b^{k,\bar{m},\eta}}, \gamma_{a^{k,\bar{m},\eta} \to b^{k,\bar{m},\eta}}$, and $\gamma_{a \to b}$ consist of the same transitions.

Proof. For (i), we have

$$\begin{aligned} c(a, a^{\bar{m},k,\rho}) &= \pi(\bar{m}, x) - \pi(k, x) + \pi(\bar{m}, x^{\bar{m},k}) - \pi(k, x^{\bar{m},k}) + \cdots + \pi(\bar{m}, x^{\bar{m},k,\rho-1}) - \pi(k, x^{\bar{m},k,\rho-1}) \\ &= \rho(\pi(\bar{m}, x) - \pi(k, x)) + \frac{\rho(\rho-1)}{2}\frac{1}{n}(-A_{\bar{m}\bar{m}} + A_{\bar{m}k} + A_{k\bar{m}} - A_{kk}). \end{aligned}$$

For (ii), first using (i) (by setting $b^{k,\bar{m},\eta} = a$), we first find that

$$c(b^{k,\bar{m},\eta}, b) - c(a^{k,\bar{m},\eta}, a) = \eta[(\pi(\bar{m}, b^{k,\bar{m},\eta}) - \pi(k, b^{k,\bar{m},\eta}) - (\pi(\bar{m}, a^{k,\bar{m},\eta}) - \pi(k, a^{k,\bar{m},\eta}))].$$

Then we have

$$\begin{aligned}&\eta[c(a,a^{\bar m,k,\rho})-c(b,b^{\bar m,k,\rho})]+\rho[c(b^{k,\bar m,\eta},b)-c(a^{k,\bar m,\eta},a)]\\=&\eta\rho[(\pi(\bar m,a)-\pi(k,a))-(\pi(\bar m,b)-\pi(k,b))]+\eta\rho[(\pi(\bar m,b^{k,\bar m,\eta})-\pi(k,b^{k,\bar m,\eta})-(\pi(\bar m,a^{k,\bar m,\eta})-\pi(k,a^{k,\bar m,\eta}))]=0\end{aligned}$$

For (iii), suppose that $(a,b)=(a_1,a_2,\cdots,a_T)$ where $a_T=b$. Then $a_{t+1}=(a_t)^{i_t,l_t}$ for some i_t,l_t. First we find

$$\begin{aligned}&\eta[c({a_t}^{\bar m,k,\rho},({a_t}^{\bar m,k,\rho})^{i_t,l_t})-c(a_t,{a_t}^{i_t,l_t})]+\rho[c({a_t}^{k,\bar m,\eta},({a_t}^{k,\bar m,\eta})^{i_t,l_t})-c(a_t,{a_t}^{i_t,l_t})]\\=&\eta[\pi(\bar m,{a_t}^{\bar m,k,\rho}-a_t)-\pi(l_t,{a_t}^{\bar m,k,\rho}-a_t)]+\rho[\pi(\bar m,{a_t}^{k,\bar m,\eta}-a_t)-\pi(l_t,{a_t}^{k,\bar m,\eta}-a_t)]\\=&\frac{1}{n}\eta[\rho(-A_{\bar m\bar m}+A_{\bar mk})-\rho(-A_{l_t\bar m}+A_{l_tk})]+\rho[\eta(-A_{\bar mk}+A_{\bar m\bar m})-\eta(-A_{l_tk}+A_{l_t\bar m})]=0\end{aligned}$$

We thus find that

$$\begin{aligned}&\eta[c(a^{\bar m,k,\rho},b^{\bar m,k,\rho})-c(a,b)]+\rho[c(a^{k,\bar m,\eta},b^{k,\bar m,\eta})-c(a,b)]\\=&\sum_{t=1}^{T-1}\eta[c({a_t}^{\bar m,k,\rho},({a_t}^{\bar m,k,\rho})^{\bar m,l_t})-c(a_t,{a_t}^{\bar m,l_t})]+\rho[c({a_t}^{k,\bar m,\eta},({a_t}^{k,\bar m,\eta})^{\bar m,l_t})-c(a_t,{a_t}^{\bar m,l_t})]=0\end{aligned}$$

□

Next, we show the following extended version of comparison principle 2, where we e denote by $(\bar m,k;\eta)$ η-times consecutive transitions from $\bar m$ to k. Also, let $x^{\bar m,k,\eta}$ be a new state induced by the agents' η-times consecutive switches from $\bar m$ to k from an old state, x.

Lemma A.2. *Consider the following paths (see Panel C, Figure 3):*

$$\begin{array}{lllllllll}\gamma & : & x & \xrightarrow[(\bar m,k;\eta)]{} & x^{\bar m,k,\eta} & \longrightarrow & y & \cdots & z & \xrightarrow[(\bar m,k;\rho)]{} & z^{\bar m,k,\rho}\\ \gamma' & : & x & \xrightarrow[(\bar m,k;\eta)]{} & x^{\bar m,k,\eta} & \xrightarrow[(\bar m,k;\rho)]{} & x^{(\bar m,k,\eta)(\bar m,k,\rho)} & \longrightarrow & y^{\bar m,k,\rho} & \cdots & z^{\bar m,k,\rho}\\ \gamma'' & : & x & \longrightarrow & y^{k,\bar m,\eta} & \cdots & z^{k,\bar m,\eta} & \xrightarrow[(\bar m,k;\eta)]{} & z & \xrightarrow[(\bar m,k;\rho)]{} & z^{\bar m,k,\rho}\end{array}$$

where $\cdots$ denotes the same transitions. Then the following holds:

$$\eta[I^{(n)}(\gamma)-I^{(n)}(\gamma')]+\rho[I^{(n)}(\gamma)-I^{(n)}(\gamma'')]=0.$$

Thus, either

$$I^{(n)}(\gamma)\geq I^{(n)}(\gamma')\quad or\quad I^{(n)}(\gamma)\geq I^{(n)}(\gamma'')$$

holds.

Proof. We find that

$$\begin{aligned}
&\eta[I(\gamma') - I(\gamma)] + \rho[I(\gamma'') - I(\gamma)] \\
=&\underbrace{\eta[c(x^{\bar{m},k,\eta}, x^{(\bar{m},k,\eta)(\bar{m},k,\rho)}) - c(z, z^{\bar{m},k,\rho})] + \rho[c(z^{k,\bar{m},\eta}, z) - c(x, x^{\bar{m},k,\eta})]}_{\text{(i)}} \\
+&\underbrace{\eta[c(x^{(\bar{m},k,\eta)(\bar{m},k,\rho)}, y^{\bar{m},k,\rho}) - c(x^{\bar{m},k,\eta}, y)] + \rho[c(x, y^{k,\bar{m},\eta}) - c(x^{\bar{m},k,\eta}, y)]}_{\text{(ii)}} \\
+&\underbrace{\eta[I(\gamma_{y^{\bar{m}k,\rho}\to z^{\bar{m},k,\rho}}) - I(\gamma_{y\to z})] + \rho[I(\gamma_{y^{k,\bar{m},\eta}\to z^{k,\bar{m},\eta}}) - I(\gamma_{y\to z})]}_{\text{(iii)}}
\end{aligned}$$

Then for (i), if we let $a = x^{\bar{m},k,\eta}$ and $b = z$ in Lemma A.1 (ii), we have (i)$= 0$. For (ii), if we let $a = x^{(\bar{m},k,\eta)}$ and $b = y$ in Lemma A.1 (ii), we have (ii)$= 0$. For (iii), if we let $a = y$ and $b = z$ in Lemma A.1 (iii), we have (iii)$= 0$. ☐

Then, **Part (ii)** follows from Lemma A.2. Suppose that $\gamma \in \mathcal{K}_{\bar{m}}$. Then, by applying Lemma A.2 repeatedly, we collect the same transitions and find $\tilde{\gamma} \in \mathcal{K}_{\bar{m}}$ such that $I(\tilde{\gamma}) \leq I(\gamma)$. Thus we obtain the desired result. ☐

Proof of Proposition 4.2. Recall that

$$D^{(n)}(e_{\bar{m}}) := \{x \in \Delta^{(n)} : \pi(\bar{m}, x) \geq \pi(l, x) \text{ for all } l\}$$

and let

$$\bar{D}(e_{\bar{m}}) := \{p \in \Delta : \pi(\bar{m}, p) \geq \pi(l, p) \text{ for all } l\} \tag{A.1}$$

and $\partial\bar{D}(e_{\bar{m}})$ be the boundary of $\bar{D}(e_{\bar{m}})$. The following lemma serves to find the continuous version of the cost function, $c(x, x^{i,j})$. Suppose that $p, q \in \Delta$ with $q = p + \alpha(e_i - e_j)$ for some $\alpha > 0$. If $p, q \in \bar{D}(e_{\bar{m}})$, we define

$$\bar{c}(p, q) := \frac{1}{2}(p_j - q_j)(\pi(\bar{m}, p + q) - \pi(i, p + q)). \tag{A.2}$$

Lemma A.3. *Let $\gamma = \gamma_{x\to y}$ be a straight-line path between $x^{(n)}$ and $y^{(n)}$ in $D(e_{\bar{m}}) \subset \Delta^{(n)}$ with $y^{(n)} = x^{(n)} + \frac{M^{(n)}}{n}(e_i - e_j)$. Suppose that $x^{(n)} \to p$ and $y^{(n)} \to q$ for $p, q \in \Delta$ as $n \to \infty$. Then,*

$$\lim_{n\to\infty} \frac{1}{n} I^{(n)}(\gamma_{x\to y}) = \frac{1}{2}(p_j - q_j)(\pi(\bar{m}, p + q) - \pi(i, p + q))$$

Proof. Since the path lies in $D(e_{\bar{m}})$ we have

$$I^{(n)}(\gamma_{x\to y}) = \sum_{\iota=0}^{M^{(n)}-1} \left[\pi\left(\bar{m}, x^{(n)} + \frac{\iota}{n}(e_i - e_j)\right) - \pi\left(i, x^{(n)} + \frac{\iota}{n}(e_i - e_j)\right)\right] . \tag{A.3}$$

Now using that $1 + 2 + \cdots + K - 1 = (K-1)K/2$, we obtain

$$\sum_{t=0}^{M^{(n)}-1} (x^{(n)} + \frac{\iota}{n}(e_i - e_j)) = M^{(n)} x^{(n)} + \frac{M^{(n)}(M^{(n)} - 1)}{2} \frac{1}{n}(e_i - e_j) = M^{(n)} \frac{x^{(n)} + y^{(n)}}{2} - \frac{M^{(n)}}{2} \frac{1}{n}(e_i - e_j) . \tag{A.4}$$

By combining equations (A.3) and (A.4) and noting that $\frac{M^{(n)}}{n} \to p_j - q_j$ as $n \to \infty$, we obtain the desired result. $\square$

The expression of costs for continuous paths in Lemma 2 in Sandholm and Staudigl (2016) is the same as the cost expression in Lemma A.3, since continuous paths in Lemma 2 in Sandholm and Staudigl (2016) belong to the special class of paths obtained by comparison principles. Next, we prove the following lemma.

Lemma A.4. *Suppose that $X^{(n)} \subset X$ and $f : X \to \mathbb{R}$ is a continuous function that admits a minimum and $f^{(n)} : X \to \mathbb{R}$. Suppose also that for all $x \in X$, there exists $\{x^{(n)}\}$ such that $x^{(n)} \in X^{(n)}$, $x^{(n)} \to x$, and $f^{(n)}(x^{(n)}) \to f(x)$. Then, we have*

$$\min_{x \in X^{(n)}} f^{(n)}(x) \to \min_{x \in X} f(x)$$

Proof. Let $\{x^{(n)}\}_n$ be the sequence of minimizers of $\min_{x\in X^{(n)}} f^{(n)}(x)$ and x^* be the minimizer of $\min_{x\in X} f(x)$. Suppose that $f^{(n)}(x^{(n)})$ does not converge to $f(x^*)$. Then there exist $\epsilon_0 > 0$ and $\{n_k\}$ such that

$$f^{(n_k)}(x^{(n_k)}) \geq f(x^*) + \epsilon_0. \tag{A.5}$$

Further, from the hypothesis, we choose $y^{(n)}$ such $y^{(n)} \to x^*$. Since $\{x^{(n)}\}$ is the sequence of minimizers, we have

$$f^{(n^k)}(y^{(n^k)}) \geq f^{(n^k)}(x^{(n^k)}) \tag{A.6}$$

Now, by taking $k \to \infty$ in equations (A.5) and (A.6), we find that $f(x^*) \geq f(x^*)+\epsilon_0$, which is a contradiction. $\square$

Now we let $X^{(n)} := \mathcal{K}^{(n)}_{\bar{m}}$ and $X = \mathcal{K}_{\bar{m}}$ and $f^{(n)} = \frac{1}{n} I^{(n)}$ and $f = \bar{I}$. Then Lemmas A.3 and A.4 show that

$$\lim_{n\to\infty} \frac{1}{n} \min\{I^{(n)}(\gamma) : \gamma \in \mathcal{K}^{(n)}_{\bar{m}}\} = \min\{\bar{I}(\zeta) : \zeta \in \mathcal{K}_{\bar{m}}\} = \min\{\omega(t) : \zeta(t) \in \mathcal{K}_{\bar{m}}\}$$

$\square$

Proof of Proposition 4.3. The proof of Proposition 4.3 follows from Lemmas A.5 and A.6.

Lemma A.5. *Let $r \in \bar{D}(e_{\bar{m}})$. Suppose that*

$$w = r + \alpha(e_k - e_{\bar{m}}), \pi(\bar{m}, w) = \pi(k, w), \text{ and } w \notin \bar{D}(e_{\bar{m}}).$$

Then there exists $j \neq k, \bar{m}$ and $\beta < \alpha$ such that

$$z := r + \beta(e_j - e_{\bar{m}}), \pi(\bar{m}, z) = \pi(j, z), \text{ and } \pi(j, r) > \pi(k, r)$$

Proof. Since $w \notin \bar{D}(e_{\bar{m}})$, there exists $j \neq k, \bar{m}$ such that $\pi(j, w) > \pi(\bar{m}, w)$. Since $\pi(\bar{m}, r) \geq \pi(j, r)$, there exists $0 < \alpha' < \alpha$ such that $\nu = r + \alpha'(e_k - e_{\bar{m}})$ and

$$\pi(\bar{m}, \nu) = \pi(j, \nu).$$

Let $o' = r + \alpha(e_j - e_{\bar{m}})$. Note that $o' = \nu - \alpha'(e_k - e_{\bar{m}}) + \alpha(e_j - e_{\bar{m}})$. Then

$$\begin{aligned}\pi(j - \bar{m}, o') &= \pi(j - \bar{m}, -\alpha'(e_k - e_{\bar{m}}) + \alpha(e_j - e_{\bar{m}})) \\ &= -\alpha'\pi(\bar{m} - j, e_{\bar{m}} - e_k) + \alpha\pi(\bar{m} - j, e_{\bar{m}} - e_j) \\ &> \alpha(\pi(j - \bar{m}, e_j - e_{\bar{m}}) - \pi(\bar{m} - j, e_{\bar{m}} - e_k)) \\ &> 0\end{aligned}$$

Thus since $\pi(\bar{m}, r) \geq \pi(j, r)$, there exists $z = r + \beta(e_j - e_{\bar{m}})$ such that $\pi(\bar{m}, z) = \pi(j, z)$ and $\beta < \alpha$. Next, we show that $\pi(j, r) > \pi(k, r)$. Suppose that $\pi(k, r) \geq \pi(j, r)$. Then we find

$$\pi(\bar{m} - j, w) = \pi(\bar{m} - j, w) - \pi(\bar{m} - k, w) = \pi(k, w) - \pi(j, w) = \pi(k - j, r) + \alpha\pi(k - j, e_k - e_{\bar{m}}) > 0$$

which is a contradiction to the fact that $\pi(\bar{m} - j, \nu) = \pi(\bar{m} - j, r + \alpha'(e_k - e_{\bar{m}})) = 0$ for $\alpha' < \alpha$. Thus, we have $\pi(j, r) > \pi(k, r)$.

□

Lemma A.6. *Let $r \in \bar{D}(e_{\bar{m}})$ and $q \in \partial\bar{D}(e_{\bar{m}})$ and $q = r + t_L(e_l - e_{\bar{m}})$. Suppose that*

$$\pi(\bar{m}, q) = \pi(k_1, q) \text{ and } \pi(\bar{m}, q) = \pi(k_2, q). \tag{A.7}$$

where $k_1 \neq k_2$. Then there exists $p \in \partial D(e_{\bar{m}})$ such that $j \neq l, \bar{m}$ and $p = r + \beta(e_j - e_{\bar{m}})$, where $0 < \beta < t_L$,

$$\pi(\bar{m}, p) = \pi(j, p) \quad \text{and} \quad c(r, p) < c(r, q).$$

Proof. From the condition, t_L is the length of transition from $\bar{m}$ to l, leading to q. Because of (A.7), we can choose $k \neq l$ such that

$$\pi(\bar{m}, q) = \pi(k, q).$$

Let $o := r + t_L(e_k - e_{\bar{m}})$. That is, o is the point obtained from r by t_L transitions from $\bar{m}$ to k). Since

$$\begin{aligned}&\pi(k - \bar{m}, r + t_L(e_k - e_{\bar{m}})) = \pi(k - \bar{m}, q + t_L(e_{\bar{m}} - e_l) + t_L(e_k - e_{\bar{m}})) \\ =&t_L\pi(k - \bar{m}, e_k - e_l) > 0\end{aligned}$$

hold from the **MBP**, we have

$$\pi(\bar{m}, r) \geq \pi(k, r) \text{ and } \pi(\bar{m}, o) < \pi(k, o)$$

and since the payoff function is linear and the game is a coordination game, there exists p such that $p = r + \alpha(e_k - e_{\bar{m}})$, where $\alpha > 0$ and $\pi(\bar{m}, p) = \pi(k, p)$. Then $o = p + (t_L - \alpha)(e_k - e_{\bar{m}})$. Thus

$$\begin{aligned}0 < \pi(k, o) - \pi(\bar{m}, o) &= \pi(k - \bar{m}, p + (t_L - \alpha)(e_k - e_{\bar{m}})) \\ &\leq (t_L - \alpha)\pi(k - \bar{m}, e_k - e_{\bar{m}})\end{aligned}$$

Thus from the **MBP**, we find $t_L > \alpha$ which implies that $p_k - r_k < q_l - r_l$. We divide cases.

Step 1. Suppose that $p \in \bar{D}(e_{\bar{m}})$. We also find

$$\begin{aligned}c(r,q)-c(r,p) =&\frac{1}{2}t_L\pi(\bar{m}-l,r+q)-\frac{1}{2}(p_k-r_k)\pi(\bar{m}-k,r+p)\\ \geq&\frac{1}{2}t_L(\pi(\bar{m}-l,r+q)-\pi(\bar{m}-k,r+p))=\frac{1}{2}t_L(\pi(k,r)-\pi(l,r))\\ =&\frac{1}{2}t_L\pi(k-l,q+t_L(e_{\bar{m}}-e_l))=\frac{1}{2}t_L\pi(\bar{m}-l,q)+\frac{1}{2}t_L^2\pi(k-l,e_{\bar{m}}-e_l)>0\end{aligned}$$

where we used $\pi(\bar{m}-l,q)\geq 0$, $\pi(k,q)=\pi(\bar{m},q)$, and the **MBP**. Thus we take $\beta:=\alpha$ and $j:=k$ and obtain the desired result.

Step 2. Suppose that $p \notin \bar{D}(e_{\bar{m}})$. We use Lemma A.5. By taking $w=p$ and using Lemma A.5, we find z. If $z \in \bar{D}(e_{\bar{m}})$, then we set $p'=z$. Otherwise, we apply the same argument using Lemma A.5 and to find z closer to r. In this way, we can find $j_1, j_2, \cdots$. Note that no two indices, j_1, j_2, are the same since if $j=j_1=j_2$ then $\pi(\bar{m}-j,r+\beta_1(e_j-e_{\bar{m}}))=\pi(\bar{m}-j_1,r+\beta_1(e_{j_1}-e_{\bar{m}}))=\pi(\bar{m}-j_2,r+\beta_2(e_{j_2}-e_{\bar{m}})=\pi(\bar{m}-j,r+\beta_2(e_j-e_{\bar{m}}))$. Thus we find $\beta_1=\beta_2$ which is a contradiction. Since the number of strategies is finite, we can find $z \in \bar{D}(e_{\bar{m}})$. Next, we show that $j \neq l$. If $j=l$, $\pi(\bar{m},z)=\pi(l,z)$. Thus, we find that

$$\begin{aligned}0 \leq&\pi(\bar{m}-l,r+t_L(e_l-e_{\bar{m}}))-\pi(\bar{m}-l,r+\beta(e_l-e_{\bar{m}}))\\ =&\pi(\bar{m}-l,(t_L-\beta)(e_l-e_{\bar{m}}))=(t_L-\beta)(-A_{\bar{m}\bar{m}}+A_{\bar{m}l}+A_{l\bar{m}}-A_{ll})\end{aligned}$$

and thus we find $t_L \leq \beta$ which is a contradiction. So we have $j \neq l$. Then observe that $p'_j - r_j < \beta < t_L$. Then, we compute as follows:

$$\begin{aligned}c(r,q)-c(r,p') =& \frac{1}{2}t_L\pi(\bar{m}-l,r+q)-\frac{1}{2}(p'_j-r_j)\pi(\bar{m}-j,r+p')\\ \geq& \frac{1}{2}t_L(\pi(\bar{m}-l,r+q)-\pi(\bar{m}-j,r+p'))=\frac{1}{2}t_L(\pi(j,r)-\pi(l,r))\\ >& \frac{1}{2}t_L(\pi(k,r)-\pi(l,r))>0\end{aligned}$$

Thus, we can take $p=p'$.

□

Now, let $t^*=((t_1,t_2,\cdots,t_L);(i_1,i_2,\cdots,i_L))$ be the solution to the minimization problem and $(\bar{m}\to i_1, \bar{m}\to i_2,\cdots,\bar{m}\to i_L)$ be the corresponding transitions. Suppose that (40) does not hold. Then there exists k_1 and k_2, $k_1 \neq k_2$, such that

$$\pi(\bar{m},q(t^*))=\pi(k_1,q(t^*)) \text{ and } \pi(\bar{m},q(t^*))=\pi(k_2,q(t^*))$$

We apply Lemma A.6 and can obtain a lower cost exit path, s^* such that $\omega(s^*)<\omega(t^*)$, which is a contradiction to optimality of t^*. □

Proof of Proposition 4.4. Suppose that $t_l^*>0$ for some $l \neq k$. To simplify notation, let $q=q(t^*)$ and

$t^* = (t_1^*, \cdots, t_K^*)$ and define

$$t_\epsilon^+ = t^* + \epsilon_k(e_k - e_{\bar{m}}) - \epsilon_l(e_l - e_{\bar{m}}), \;\; t_\epsilon^- = t^* - \epsilon_k(e_k - e_{\bar{m}}) + \epsilon_l(e_l - e_{\bar{m}})$$

Then, we have

$$\begin{aligned}\pi(\bar{m}, q(t_\epsilon^+)) - \pi(k, q(t_\epsilon^+)) =& \epsilon_k \pi(\bar{m}, k - \bar{m}) - \epsilon_l \pi(\bar{m}, l - \bar{m}) - \epsilon_k \pi(k, k - \bar{m}) + \epsilon_l \pi(k, l - \bar{m}) \\ =& - \epsilon_k(A_{\bar{m}\bar{m}} - A_{\bar{m}k} + A_{kk} - A_{k\bar{m}}) + \epsilon_l(A_{\bar{m}\bar{m}} - A_{k\bar{m}} + A_{\bar{m}l} - A_{kl}) \\ \pi(\bar{m}, q(t_\epsilon^-)) - \pi(k, q(t_\epsilon^-)) =& \epsilon_k(A_{\bar{m}\bar{m}} - A_{\bar{m}k} + A_{kk} - A_{k\bar{m}}) - \epsilon_l(A_{\bar{m}\bar{m}} - A_{k\bar{m}} + A_{\bar{m}l} - A_{kl})\end{aligned}$$

and similarly, for $j \neq k$, we find that

$$\begin{aligned}\pi(\bar{m}, q(t_\epsilon^+)) - \pi(j, q(t_\epsilon^+)) =& \pi(\bar{m}, q) - \pi(j, q) \\ & + \epsilon_k \pi(\bar{m}, k - \bar{m}) - \epsilon_l \pi(\bar{m}, l - \bar{m}) - \epsilon_k \pi(j, k - \bar{m}) + \epsilon_l \pi(j, l - \bar{m}) \\ =& \pi(\bar{m}, q) - \pi(j, q) \\ & - \epsilon_k(A_{\bar{m}\bar{m}} - A_{\bar{m}k} + A_{jk} - A_{j\bar{m}}) + \epsilon_l(A_{\bar{m}\bar{m}} - A_{j\bar{m}} + A_{\bar{m}l} - A_{jl}) \\ \pi(\bar{m}, q(t_\epsilon^-)) - \pi(j, q(t_\epsilon^-)) =& \pi(\bar{m}, q) - \pi(j, q) \\ & \epsilon_k(A_{\bar{m}\bar{m}} - A_{\bar{m}k} + A_{jk} - A_{j\bar{m}}) - \epsilon_l(A_{\bar{m}\bar{m}} - A_{j\bar{m}} + A_{\bar{m}l} - A_{jl})\end{aligned}$$

Thus, we can choose small $\epsilon_k, \epsilon_l > 0$ such that

$$\begin{aligned}&\pi(\bar{m}, q(t_\epsilon^+)) = \pi(k, q(t_\epsilon^+)), \text{ and } \pi(\bar{m}, q(t_\epsilon^+)) > \pi(j, q(t_\epsilon^+)) \text{ for all } l \neq k \\ &\pi(\bar{m}, q(t_\epsilon^-)) = \pi(k, q(t_\epsilon^-)), \text{ and } \pi(\bar{m}, q(t_\epsilon^-)) > \pi(j, q(t_\epsilon^-)) \text{ for all } l \neq k,\end{aligned}$$

which show that t_ϵ^+ and t_ϵ^- both satisfy the constraints. Recall

$$H_{i,j:k} := (A_{ii} - A_{ji}) - (A_{ik} - A_{jk}).$$

Then we find that

$$\begin{aligned}&\text{If } t_l \text{ is ahead of } t_k, \; (\omega(t_\epsilon^+) - \omega(t)) - (\omega(t) - \omega(t_\epsilon^-)) = -\epsilon_l^2 \pi(\bar{m} - l, \bar{m} - l) + 2\epsilon_l \epsilon_k \pi(\bar{m} - k, \bar{m} - l) - \epsilon_k^2 \pi(\bar{m} - k, \bar{m} - k) \\ &\text{If } t_k \text{ is ahead of } t_l, \; (\omega(t_\epsilon^+) - \omega(t)) - (\omega(t) - \omega(t_\epsilon^-)) = -\epsilon_l^2 \pi(\bar{m} - l, \bar{m} - l) + 2\epsilon_l \epsilon_k \pi(\bar{m} - l, \bar{m} - k) - \epsilon_k^2 \pi(\bar{m} - k, \bar{m} - k)\end{aligned}$$

Thus, we find that

$$\begin{aligned}&(\omega(t_\epsilon^+) - \omega(t)) - (\omega(t) - \omega(t_\epsilon^-)) = -H_{\bar{m}k:k}\epsilon_k^2 + 2\max\{H_{\bar{m}k:l}, H_{\bar{m}l:k}\}\epsilon_k\epsilon_l - H_{\bar{m}l:l}\epsilon_l^2 \\ &\leq -H_{\bar{m}k:k}\epsilon_k^2 + 2\sqrt{H_{\bar{m}k:k}}\sqrt{H_{\bar{m}l:l}}\epsilon_k\epsilon_l - H_{\bar{m}l:l}\epsilon_l^2 \leq -(\sqrt{H_{\bar{m}k:k}}\epsilon_k - \sqrt{H_{\bar{m}l:l}}\epsilon_l)^2 < 0\end{aligned}$$

where we use

$$\max\{H_{\bar{m}k:l}, H_{\bar{m}l:k}\} < H_{\bar{m}k:k}, \;\; \max\{H_{\bar{m}k:l}, H_{\bar{m}l:k}\} < H_{\bar{m}l:l}.$$

from **MBP**. This shows that either $\omega(t_\epsilon^+) < \omega(t)$ or $\omega(t) > \omega(t_\epsilon^-)$ holds, a contradiction to the optimality of t. □

Proof of Theorem 4.1. Let t^* be the solution to the minimization problem:

$$\min\{\omega(t) : \zeta(t) \in \bar{\mathcal{K}}_{\bar{m}}\}.$$

Propositions 4.4 and 4.3 show that there exists k such that $t^*_k > 0$ and $t^*_l = 0$ for all $l \neq k$ and Theorem 4.1 follows immediately from this and Proposition 4.2. □

B. Exit problem: two-population models

The following lemma is analogous to Lemma 4.1, which shows that it always costs less (or the same) to first switch from strategy $\bar{m}$, than from other strategies.

Lemma B.1. *Suppose that the **WBP** holds.*

$$c^{(n)}(x^{\beta,\bar{m},k}, x^{(\beta,\bar{m},k)(\alpha,j,h)}) - c^{(n)}(x^{\beta,i,k}, x^{(\beta,i,k)(\alpha,j,h)}) = -A^{\alpha}_{\bar{m}\bar{m}} + A^{\alpha}_{h\bar{m}} + A^{\alpha}_{\bar{m}i} - A^{\alpha}_{hi} \leq 0$$
$$c^{(n)}(x^{\alpha,\bar{m},k}, x^{(\alpha,\bar{m},k)(\beta,j,h)}) - c^{(n)}(x^{\alpha,i,k}, x^{(\alpha,i,k)(\beta,j,h)}) = -A^{\beta}_{\bar{m}\bar{m}} + A^{\beta}_{\bar{m}h} + A^{\beta}_{i\bar{m}} - A^{\beta}_{ih} \leq 0.$$

Proof. These are immediate from the definition. □

Proposition B.1 shows that Lemma B.1 can be extended to arbitrary paths. We use Proposition B.1 to show how to remove the transitions from $i \neq \bar{m}$ in a given path to achieve a lower cost. In Proposition B.1, (β, i, k), for example, refers to a transition by a β-agent from strategy i to k.

Proposition B.1. *Suppose that the **WBP** holds. We consider two paths:*

$$\gamma_1 : x \xrightarrow[(\beta,i,k)]{} x^{(1)} \xrightarrow[(\alpha,j_1,k_1)]{} x^{(2)} \xrightarrow[(\alpha,j_2,k_2)]{} x^{(3)} \cdots x^{(L-1)} \xrightarrow[(\alpha,j_L,k_L)]{} x^{(L)} \xrightarrow[(\beta,\bar{m},l)]{} y$$
$$\gamma_2 : x \xrightarrow[(\beta,\bar{m},k)]{} y^{(1)} \xrightarrow[(\alpha,j_1,k_1)]{} y^{(2)} \xrightarrow[(\alpha,j_2,k_2)]{} y^{(3)} \cdots y^{(L-1)} \xrightarrow[(\alpha,j_L,k_L)]{} y^{(L)} \xrightarrow[(\beta,i,l)]{} y$$

Then, we have $I^{(n)}(\gamma_1) \geq I^{(n)}(\gamma_2)$ and a similar statement holds for a path with transitions of α agents from i to k and $\bar{m}$ to l and transitions of α agents from $\bar{m}$ to k and from i to l.

Proof. We find that

$$\begin{aligned}
I^{(n)}(\gamma_1) &= c^{(n)}(x, x^{\beta,i,k}) + c^{(n)}(x^{\beta,i,k}, x^{(\beta,i,k)(\alpha,j_1,k_1)}) + c^{(n)}(x^{\beta,i,k}, x^{(\beta,i,k)(\alpha,j_2,k_2)}) \\
&\quad + \cdots c^{(n)}(x^{\beta,i,k}, x^{(\beta,i,k)(\alpha,j_L,k_L)}) + c^{(n)}(x^{(L)}, (x^{(L)})^{(\beta,\bar{m},l)}). \\
I^{(n)}(\gamma_2) &= c^{(n)}(x, x^{\beta,\bar{m},k}) + c^{(n)}(x^{\beta,\bar{m},k}, x^{(\beta,\bar{m},k)(\alpha,j_1,k_1)}) + c^{(n)}(x^{\beta,\bar{m},k}, x^{(\beta,\bar{m},k)(\alpha,j_2,k_2)}) \\
&\quad + \cdots c^{(n)}(x^{\beta,\bar{m},k}, x^{(\beta,\bar{m},k)(\alpha,j_L,k_L)}) + c^{(n)}(x^{(L)}, (x^{(L)})^{(\beta,i,l)})
\end{aligned}$$

from the fact that $c^{(n)}(x^{(l)}, (x^{(l)})^{\alpha,j_l,k_l}) = c^{(n)}(x^{\beta,i,k}, x^{(\beta,i,k)(\alpha,j_l,k_l)})$ for $l = 2, \cdots, L-1$ and $c(y^{(l)}, (y^{(l)})^{\alpha,j_l,k_l}) = c^{(n)}(y^{\beta,\bar{m},k}, x^{(\beta,\bar{m},k)(\alpha,j_l,k_l)})$ for $l = 2, \cdots, L-1$ (see Lemma B.2). Observe that $c^{(n)}(x, x^{\beta,\bar{m},k}) = c^{(n)}(x, x^{\beta,i,k})$ and $c^{(n)}(x^{(L)}, (x^{(L)})^{(\beta,i,l)}) = c^{(n)}(x^{(L)}, (x^{(L)})^{(\beta,\bar{m},l)})$. Then by applying Lemma 2 successively, we obtain the desired result. □

We can also collect the same transitions as follows, analogously to Proposition A.2. We also denote by $(\beta, \bar{m}, k; \eta)$ the consecutive transitions of β-agent from $\bar{m}$ to k η-times.

Proposition B.2. *Consider the following paths:*

$$\begin{array}{lllllllll}
\gamma & : & x & \xrightarrow[(\beta,\bar{m},k;\eta)]{} & x^{\beta,\bar{m},k;\eta} & \longrightarrow & y & \cdots & z & \xrightarrow[(\beta,\bar{m},k;\rho)]{} & z^{\beta,\bar{m},k;\rho} \\
\gamma' & : & x & \xrightarrow[(\beta,\bar{m},k;\eta)]{} & x^{\beta,\bar{m},k;\eta} & \xrightarrow[(\beta,\bar{m},k;\rho)]{} & x^{(\beta,\bar{m},k;\eta)(\beta,\bar{m},k;\rho)} & \longrightarrow & y^{\beta,\bar{m},k;\rho} & \cdots & z^{\beta,\bar{m},k;\rho} \\
\gamma'' & : & x & \longrightarrow & y^{\beta,k,\bar{m};\eta} & \cdots & z^{\beta,k,\bar{m};\eta} & \xrightarrow[(\beta,\bar{m},k;\eta)]{} & z & \xrightarrow[(\beta,\bar{m},k;\rho)]{} & z^{\beta,\bar{m},k;\rho}
\end{array}$$

where $\cdots$ denotes the same transitions. Then either

$$I^{(n)}(\gamma) \geq I^{(n)}(\gamma'), \quad or\ I^{(n)}(\gamma) \geq I^{(n)}(\gamma'')$$

holds. A similar statement holds for a path involving transitions of α agents' transitions.

Proof. We start with the following lemma.

Lemma B.2. *We have the following results:*

$$c^{(n)}(x, x^{\alpha,i,j}) = c^{(n)}(z, z^{\alpha,i,j}) \text{ for all } x_\beta = z_\beta$$
$$c^{(n)}(x, x^{\beta,i,j}) = c^{(n)}(z, z^{\beta,i,j}) \text{ for all } x_\alpha = z_\alpha$$

Proof. This is immediate from the definition. ☐

Next we show the following lemma.

Lemma B.3. *We have the following results:*

$$\eta[c^{(n)}(a^{\beta,\bar{m},k,\rho}, b^{\beta,\bar{m},k,\rho}) - c^{(n)}(a,b)] + \rho[c^{(n)}(a^{\beta,k,\bar{m},\eta}, b^{\beta,k,\bar{m},\eta}) - c^{(n)}(a,b)] = 0$$

Proof. Suppose that $(a,b) = (a_1, a_2, \cdots, a_T)$ where $a_T = b$. Suppose that $a_{t+1} = (a_t)^{\beta,i_t,l_t}$. Then by applying Lemma B.2, we obtain

$$\eta[c^{(n)}(a_t^{\beta,\bar{m},k,\rho}, (a_t^{\beta,\bar{m},k,\rho})^{\beta,i_t,l_t}) - c^{(n)}(a_t, a_t^{\beta,i_t,l_t})] + \rho[c^{(n)}(a_t^{\beta,k,\bar{m},\eta}, (a_t^{\beta,k,\bar{m},\eta})^{\beta,i_t,l_t}) - c^{(n)}(a_t, a_t^{\beta,i_t,l_t})] = 0$$

We next suppose that $a_{t+1} = (a_t)^{\alpha,i_t,l_t}$.

$$\begin{aligned}
&\eta[c^{(n)}(a_t^{\beta,\bar{m},k,\rho}, (a_t^{\beta,\bar{m},k,\rho})^{\alpha,i_t,l_t}) - c^{(n)}(a_t, a_t^{\alpha,i_t,l_t})] + \rho[c^{(n)}(a_t^{\beta,k,\bar{m},\eta}, (a_t^{\beta,k,\bar{m},\eta})^{\alpha,i_t,l_t}) - c^{(n)}(a_t, a_t^{\alpha,i_t,l_t}) \\
=&\eta[\pi_\alpha(\bar{m}, a_t^{\beta,\bar{m},k,\rho}) - \pi_\alpha(l_t, a_t^{\beta,\bar{m},k,\rho}) - \pi_\alpha(\bar{m}, a_t) + \pi_\alpha(l_t, a_t)] \\
&+ \rho[\pi_\alpha(\bar{m}, a_t^{\beta,k,\bar{m},\eta}) - \pi_\alpha(l_t, a_t^{\beta,k,\bar{m},\eta}) - \pi_\alpha(\bar{m}, a_t) + \pi_\alpha(l_t, a_t)] \\
=&0
\end{aligned}$$

Thus we find

$$
\begin{aligned}
&\eta[I^{(n)}(\gamma_{a^{\beta,\bar{m},k,\rho}\to b^{\beta,\bar{m},k,\rho}}) - I^{(n)}(\gamma_{a\to b})] + \rho[I^{(n)}(\gamma_{a^{\beta,k,\bar{m},\eta}\to b^{\beta,k,\bar{m},\eta}}) - I(\gamma_{a\to b})] \\
=&\sum_{t=1}^{T-1} \eta[c^{(n)}({a_t}^{\beta,\bar{m},k,\rho}, ({a_t}^{\beta,\bar{m},k,\rho})^{i_t,l_t}) - c^{(n)}(a_t, {a_t}^{\beta,i_t,l_t})] + \rho[c^{(n)}({a_t}^{\beta,k,\bar{m},\eta}, ({a_t}^{\beta,k,\bar{m},\eta})^{i_t,l_t}) - c^{(n)}(a_t, {a_t}^{i_t,l_t})] \\
=&0
\end{aligned}
$$

□

Lemma B.4. *We have the following results:*
(i) $\eta[c^{(n)}(x^{\beta,\bar{m},k;\eta}, x^{(\beta,\bar{m},k;\eta),(\beta,\bar{m},k;\rho)}) - c^{(n)}(z, z^{(\beta,\bar{m},k;\rho)})] + \rho[c^{(n)}(z^{(\beta,k,\bar{m};\eta)}, z) - c^{(n)}(x, x^{\beta,\bar{m},k;\eta})] = 0$
(ii) $\eta[c^{(n)}(x^{(\beta,\bar{m},k;\eta),(\beta,\bar{m},k;\rho)}, y^{\beta,\bar{m},k;\rho}) - c^{(n)}(x^{\beta,\bar{m},k;\eta}, y)] + \rho[c^{(n)}(x, y^{\beta,k,\bar{m};\eta}) - c^{(n)}(x^{\beta,\bar{m},k,\eta}, y)] = 0$
(iii) $\eta[I^{(n)}(\gamma_{y^{\beta,\bar{m},k;\rho}\to z^{\beta,\bar{m},k\rho}}) - I^{(n)}(\gamma_{y\to z})] + \rho[I^{(n)}(\gamma_{y^{\beta,k,\bar{m},\eta}\to z^{\beta,k,\bar{m},\eta}}) - I^{(n)}(\gamma_{y\to z})] = 0$

Proof. (i) By applying Lemma B.2, we find that

$$
\begin{aligned}
&\eta[c^{(n)}(x^{\beta,\bar{m},k;\eta}, x^{(\beta,\bar{m},k;\eta),(\beta,\bar{m},k;\rho)}) - c^{(n)}(z, z^{(\beta,\bar{m},k;\rho)})] + \rho[c^{(n)}(z^{(\beta,k,\bar{m};\eta)}, z) - c^{(n)}(x, x^{\beta\bar{m},k;\eta})] \\
=&\eta[c^{(n)}(x, x^{(\beta,\bar{m},k;\rho)} - c^{(n)}(z, z^{(\beta,\bar{m},k;\rho)})] + \rho[c^{(n)}(z, z^{(\beta,\bar{m},k;\eta)}) - c^{(n)}(x, x^{\beta,\bar{m},k;\eta})] \\
=&\eta\rho[\pi_\beta(\bar{m}, x) - \pi_\beta(k, x) - \pi_\beta(\bar{m}, z) + \pi_\beta(k, z)] + \rho\eta[\pi_\beta(\bar{m}, z) - \pi_\beta(k, z) - \pi_\beta(\bar{m}, x) + \pi_\beta(k, x)] \\
=&0
\end{aligned}
$$

(ii) follows from by letting $a := x^{\beta,\bar{m},k;\eta}$ and $b := y$ in Lemma B.4 and (iii) follows from by letting $a := y$ and $b := z$ in Lemma B.4. □

Proof of Proposition B.2. We find that

$$
\begin{aligned}
&\eta[I^{(n)}(\gamma') - I^{(n)}(\gamma)] + \rho[I^{(n)}(\gamma'') - I^{(n)}(\gamma)] \\
=&\underbrace{\eta[c^{(n)}(x^{\beta,\bar{m},k;\eta}, x^{(\beta,\bar{m},k;\eta),(\beta,\bar{m},k;\rho)}) - c^{(n)}(z, z^{(\beta,\bar{m},k;\rho)})] + \rho[c^{(n)}(z^{(\beta,k,\bar{m};\eta)}, z) - c^{(n)}(x, x^{\beta,\bar{m},k;\eta})]}_{\text{(i)}} \\
+&\underbrace{\eta[c^{(n)}(x^{(\beta,\bar{m},k;\eta),(\beta,\bar{m},k;\rho)}, y^{\beta,\bar{m},k;\rho}) - c^{(n)}(x^{\beta,\bar{m},k;\eta}, y)] + \rho[c^{(n)}(x, y^{\beta,k,\bar{m};\eta}) - c^{(n)}(x^{\beta,\bar{m},k,\eta}, y)]}_{\text{(ii)}} \\
+&\underbrace{\eta[I^{(n)}(\gamma_{y^{\beta,\bar{m},k;\rho}\to z^{\beta,\bar{m},k\rho}}) - I^{(n)}(\gamma_{y\to z})] + \rho[I^{(n)}(\gamma_{y^{\beta,k,\bar{m},\eta}\to z^{\beta,k,\bar{m},\eta}}) - I^{(n)}(\gamma_{y\to z})]}_{\text{(iii)}}
\end{aligned}
$$

and Lemma B.4 (i), (ii), and (iii) show the desired result. □

We also define $\mathcal{J}_{\bar{m}}^{(n)}$ and $\mathcal{K}_{\bar{m}}^{(n)}$ analogously to equations (32) and (34). That is, $\mathcal{J}_{\bar{m}}^{(n)}$ is the set of all paths in which all the transitions are from strategy $\bar{m}$ and $\mathcal{K}_{\bar{m}}^{(n)}$ is the set of all paths consisting of consecutive transitions from $\bar{m}$ to some other strategy. From Propositions B.1 and B.2, we next show that the minimum transition cost path γ involves only transitions from $\bar{m}$.

Proposition B.3. *Suppose that the **WBP** holds.*

(i) We have

$$\min\{I^{(n)}(\gamma) : \gamma \in \mathcal{G}^{(n)}_{\bar{m}}\} = \min\{I^{(n)}(\gamma) : \gamma \in \mathcal{J}^{(n)}_{\bar{m}}\}.$$

(ii) We have

$$\min\{I^{(n)}(\gamma) : \gamma \in \mathcal{G}^{(n)}_{\bar{m}}\} = \min\{I^{(n)}(\gamma) : \gamma \in \mathcal{K}^{(n)}_{\bar{m}}\}.$$

Proof. For the proof, we suppress the superscript (n). Part (i). Let $\gamma \in \mathcal{G}_{\bar{m}} \backslash \mathcal{J}_{\bar{m}}$. Let the last transition of γ be from z to $z^{\beta,i,l}$ for some $i \neq \bar{m}$. Since $c^{(n)}(z, z^{\beta,i,l}) = c^{(n)}(z, z^{\beta,\bar{m},l})$, by modifying the last transition from $z^{\beta,i,l}$ to $z^{\beta,\bar{m},l}$ the cost will not be changed. Now, suppose that x is the last state from which a transition occurs from $i \neq \bar{m}$ in the modified path (see γ_1 in Proposition B.1). Then, by applying Proposition B.1, we obtain the new path whose last transition is from $i \neq \bar{m}$ (see γ_2 in Proposition B.1). By changing this last transition again, we can obtain a new modified path. In this way, we can remove all β-agents' transitions from $i \neq \bar{m}$. Similarly, we can also remove all α-agents' transitions from $i \neq \bar{m}$ using the corresponding part for α agents in Proposition B.1. Thus, we can obtain the desired results. Part (ii) immediately follows from Proposition B.2. □

Next, we consider the continuous limit. For this, we define a cost function $\bar{c}(\mathbf{p}, \mathbf{q})$, for $\mathbf{p} = (p_\alpha, p_\beta), \mathbf{q} = (q_\alpha, q_\beta) \in \Delta_\alpha \times \Delta_\beta$. Let $\mathbf{q} = \mathbf{p} + (\rho(e^\alpha_i - e^\alpha_j), 0)$ or $\mathbf{q} = \mathbf{p} + (0, \rho(e^\beta_i - e^\beta_j))$ for some $\rho > 0$. If $\mathbf{p}, \mathbf{q} \in \bar{D}(e_{\bar{m}})$,

$$\bar{c}(\mathbf{p}, \mathbf{q}) = (p_{\alpha,j} - q_{\alpha,j})(\pi_\alpha(\bar{m}, p) - \pi_\alpha(j, p)) \text{ or } \bar{c}(\mathbf{p}, \mathbf{q}) = (p_{\beta,j} - q_{\beta,j})(\pi_\beta(\bar{m}, p) - \pi_\beta(j, p)).$$

We similarly define $\bar{\mathcal{K}}_{\bar{m}}$ as in the one population model and from $\zeta = \zeta(t) \in \mathcal{K}_{\bar{m}}$, where $t = ((t^\alpha, t^\beta); (i^\alpha, j^\beta)) = (((t^\alpha_1, \cdots, t^\alpha_K), (t^\beta_1, \cdots, t^\beta_K)); (i^\alpha_1, \cdots, i^\alpha_K); (j^\beta_1, \cdots, j^\beta_K))$ and define $\omega(t) = \sum_{s=0}^{K-1} \bar{c}(\mathbf{p}^{(s)}, \mathbf{p}^{(s+1)})$. Then, we have the following lemma.

Lemma B.5. *Let $\bar{t}^\beta$, i^α, j^α be fixed. Then $\omega(\cdot, \bar{t}^\beta)$ is affine. A similar statement holds for the case where $\bar{t}^\alpha$ is fixed.*

Proof. Suppose that t^α_i is associated with α agents' transitions from $\bar{m}$ to i. Similarly, t^β_j is associated with β agents' transitions from $\bar{m}$ to j. Let $\mathbf{p}$ be the state from which the transitions represented by t^α_i start. Then we find that

$$\begin{aligned}\frac{\partial \omega}{\partial t^\alpha_i} = (\pi_\alpha(\bar{m}, p_\beta) - \pi_\alpha(i, p_\beta)) &+ \bar{t}^\beta_i(-A^\beta_{ii} + A^\beta_{i\bar{m}} - A^\beta_{\bar{m}\bar{m}} + A^\beta_{\bar{m}i}) \\ &+ \sum_{j \neq i} \bar{t}^\beta_j(-A^\beta_{ij} + A^\beta_{i\bar{m}} - A^\beta_{\bar{m}\bar{m}} + A^\beta_{\bar{m}j})\end{aligned}$$

and observe that $\pi_\alpha(\bar{m}, p_\beta) - \pi_\alpha(i, p_\beta)$ depends only on $\bar{t}^\beta$; this shows that $\omega(\cdot, \bar{t}^\beta)$ is affine. □

Thus, we similarly consider

$$\min\{\omega(t) : \zeta(t) \in \mathcal{K}_{\bar{m}}\}.$$

Using the characterization that ω is affine, we show that if $t^{\alpha^*}_i > 0$ in an optimal path, then $\pi_\beta(\bar{m}, \mathbf{q}^*(t^*)) = \pi_\beta(i, \mathbf{q}^*(t^*))$ at the exit point $\mathbf{q}^*(t^*)$, where $t^{\alpha^*}_i$ denotes the transition by an α-agent from strategy $\bar{m}$ to i.

Proposition B.4. *Suppose that Condition **B** holds. Then, there exists* $\zeta(t^*) \in \mathcal{K}_{\bar{m}}$ *such that* $\omega(t^*) = \min\{\omega(t) : \zeta(t) \in \mathcal{K}_{\bar{m}}\}$ *and if* $t_i^{\alpha^*} > 0$*, then* $\pi_\beta(\bar{m}, \mathbf{q}(t^*)) = \pi_\beta(i, \mathbf{q}(t^*))$ *and if* $t_j^{\beta^*} > 0$*, then* $\pi_\alpha(\bar{m}, \mathbf{q}(t^*)) = \pi_\alpha(j, \mathbf{q}(t^*))$*, where* $\mathbf{q}(t^*)$ *is the end state of* $\zeta(t^*)$*.*

Proof. Let t^* be given such that $\omega(t^*) = \min\{\omega(t) : \zeta(t) \in \mathcal{K}_{\bar{m}}\}$. Suppose that $t_i^{\alpha^*} > 0$. The other case follows similarly. Let $\bar{t}_i^\alpha$ such that

$$\pi_\beta(\bar{m}, (1-\bar{t}_i^\alpha)e_{\bar{m}}^\alpha + \bar{t}_i^\alpha e_i^\alpha) = \pi_\beta(i, (1-\bar{t}_i^\alpha)e_{\bar{m}}^\alpha + \bar{t}_i^\alpha e_{\bar{m}}^\alpha)$$

Then, we have

$$\begin{aligned}
&\pi_\beta(i, q_\alpha(t^{\alpha^*})) - \pi_\beta(\bar{m}, q_\alpha(t^{\alpha^*})) \\
=&\pi_\beta(i-\bar{m}, (1-t_i^{\alpha^*})e_{\bar{m}}^\alpha + t_i^{\alpha^*} e_i^\alpha) + \sum_{l\neq i} t_l^{\alpha^*}\pi_\beta(i-\bar{m}, e_l^\alpha - e_{\bar{m}}^\alpha) \\
=&\pi_\beta(i-\bar{m}, (1-t_i^{\alpha^*})e_{\bar{m}}^\alpha + t_i^{\alpha^*} e_i^\alpha) + \sum_{l\neq i} t_l^{\alpha^*}(A_{li}^\beta - A_{l\bar{m}}^\beta - A_{\bar{m}i}^\beta + A_{\bar{m}\bar{m}}^\beta). \qquad \text{(B.1)}
\end{aligned}$$

Now, we have two cases:

Case 1: $t_i^{\alpha^*} = \bar{t}_i^\alpha$.
Since $t^* \in \mathcal{K}_{\bar{m}}$, $\pi_\beta(i, q_\alpha(t^{\alpha^*})) - \pi_\beta(\bar{m}, q_\alpha(t^{\alpha^*})) \leq 0$, the second term in (B.1) ($\sum_{l\neq i} t_l^{\alpha^*}(A_{li}^\beta - A_{l\bar{m}}^\beta - A_{\bar{m}i}^\beta + A_{\bar{m}\bar{m}}^\beta)$) is non-positive. Also, the **WBP** implies that the same term is non-negative, and hence zero. Thus, we have $\pi_\beta(i, q_\alpha(t^{\alpha^*})) = \pi_\beta(\bar{m}, q_\alpha(t^{\alpha^*}))$, which is the desired result.

Case 2: $0 < t_i^{\alpha^*} < \bar{t}_i^\alpha$.
Suppose that

$$\pi_\beta(\bar{m}, q_\alpha(t^{\alpha^*})) > \pi_\beta(i, q_\alpha(t^{\alpha^*})). \qquad \text{(B.2)}$$

and

$$\pi_\beta(\bar{m}, q_\alpha(t^{\alpha^*})) = \pi_\beta(j_1, q_\alpha(t^{\alpha^*})), \pi_\beta(\bar{m}, q_\alpha(t^{\alpha^*})) = \pi_\beta(j_2, q_\alpha(t^{\alpha^*})), \cdots, \pi_\beta(\bar{m}, q_\alpha(t^{\alpha^*})) = \pi_\beta(j_L, q_\alpha(t^{\alpha^*})), \qquad \text{(B.3)}$$

where the other constraints for π_β are non-binding. To reach $q_\alpha(t^{\alpha^*})$, there are transitions, $\bar{m} \to j_1, \bar{m} \to j_2, \cdots, \bar{m} \to j_L$ and thus

$$q_\alpha(t^{\alpha^*}) = e_{\bar{m}}^\alpha + \sum_{l=1}^{L} t_{j_l}(e_{j_l}^\alpha - e_{\bar{m}}^\alpha) + t_i^{\alpha^*}(e_i^\alpha - e_{\bar{m}}^\alpha) + \sum_k s_k(e_k^\alpha - e_{\bar{m}})$$

And we find that

$$\pi_\beta(j_1-\bar{m},q_\alpha(t^{\alpha^*})=\sum_{l=1}^{L}\pi_\beta(j_1-\bar{m},e^\alpha_{j_l}-e^\alpha_{\bar{m}})t_{j_l}+\pi(j_1-\bar{m},e^\alpha_i-e^\alpha_{\bar{m}})t^{\alpha^*}_i+\pi(j_1-\bar{m},e^\alpha_{\bar{m}}+\sum_k s_k(e^\alpha_k-e_{\bar{m}}))$$
$$=\sum_{l=1}^{L}(A^\beta_{\bar{m}\bar{m}}-A^\beta_{j_1\bar{m}})-(A_{\bar{m}j_l}-A_{j_1j_l})t_{j_l}+\pi(j_1-\bar{m},e^\alpha_i-e^\alpha_{\bar{m}})t^{\alpha^*}_i+\pi(j_1-\bar{m},e^\alpha_{\bar{m}}+\sum_k s_k(e^\alpha_k-e_{\bar{m}}))$$
$$\cdots$$
$$\pi_\beta(j_L-\bar{m},q_\alpha(t^{\alpha^*})=\sum_{l=1}^{L}\pi_\beta(j_L-\bar{m},e^\alpha_{j_l}-e^\alpha_{\bar{m}})t_{j_l}+\pi(j_L-\bar{m},e^\alpha_i-e^\alpha_{\bar{m}})t^{\alpha^*}_i+\pi(j_L-\bar{m},e^\alpha_{\bar{m}}+\sum_k s_k(e^\alpha_k-e_{\bar{m}}))$$
$$=\sum_{l=1}^{L}(A^\beta_{\bar{m}\bar{m}}-A^\beta_{j_L\bar{m}})-(A_{\bar{m}j_l}-A_{j_Lj_l})t_{j_l}+\pi(j_L-\bar{m},e^\alpha_i-e^\alpha_{\bar{m}})t^{\alpha^*}_i+\pi(j_L-\bar{m},e^\alpha_{\bar{m}}+\sum_k s_k(e^\alpha_k-e_{\bar{m}}))$$

Thus we can regard equations in (B.3) as a set of linear equations in variables, $t_{j_1}, t_{j_2}, \cdots, t_{j_L}$. Then, from the implicit function theorem and Lemma B.6 (Condition B) we can find functions $t^*_{j_1}(t_i)$, $t^*_{j_2}(t_i), \cdots$, $t^*_{j_L}(t_i)$ satisfying (B.2) and (B.3) for all $t_i \in [t^{\alpha^*}_i-\epsilon, t^{\alpha^*}_i+\epsilon]$ for some $\epsilon > 0$. Observe that $t^*_{j_1}(t_i)$, $t^*_{j_2}(t_i), \cdots$, $t^*_{j_L}(t_i)$ are affine in t_i. Then, we define $\phi(t_i)=\omega((t_i, t^*_{j_1}(t_i), t^*_{j_2}(t_i), \cdots, t^*_{j_L}(t_i), \bar{t}_{i_1}, \bar{t}_{i_2}, \cdots, \bar{t}_{i_{L'}}), \bar{t}^\beta)$. From Lemma B.5, we see that $\phi(t_i)$ is affine with respect to t_i. We then find ϕ' and again have two cases.
Case 2-1. Suppose that $\phi'=0$. Then, by increasing t_i up to $\pi_\beta(\bar{m},\mathbf{q}(t_\alpha))=\pi_\beta(i,\mathbf{q}(t_\alpha))$, we can find t^{**} which satisfies $\omega(t^{**})=\omega(t^*)$ and obtain the desired properties in the proposition.
Case 2-2. Suppose that $\phi'\neq 0$. Then, we have either $\phi(t^{\alpha^*}_i-\epsilon)>\phi(t^{\alpha^*}_i)>\phi(t^{\alpha^*}_i+\epsilon)$ or $\phi(t^{\alpha^*}_i-\epsilon)<\phi(t^{\alpha^*}_i)<\phi(t^{\alpha^*}_i+\epsilon)$, in contradiction to the optimality of t^*. □

Lemma B.6. *The following statement holds:*

$$\pi_\kappa(\bar{m},r)=\pi_\kappa(i_1,r),\cdots,\pi_\kappa(\bar{m},r)=\pi_\kappa(i_K,r),\ r_{\bar{m}}+\sum_{i=1}^{K}r_{i_l}=1,\ \Sigma_r=\{\bar{m},i_1,\cdots,i_K\}$$

have a unique solution.
$\iff$ *$det(D)\neq 0$ where*

$$D=\begin{pmatrix} A^\kappa_{\bar{m}\bar{m}}-A^\kappa_{i_1\bar{m}}-(A^\kappa_{\bar{m}i_1}-A^\kappa i_1i_1) & \cdots & A^\kappa_{\bar{m}\bar{m}}-A^\kappa_{i_1\bar{m}}-(A^\kappa_{\bar{m}i_K}-A^\kappa_{i_1i_K}) \\ A^\kappa_{\bar{m}\bar{m}}-A^\kappa_{i_2\bar{m}}-(A^\kappa_{\bar{m}i_1}-A^\kappa_{i_2i_1}) & \cdots & A^\kappa_{\bar{m}\bar{m}}-A^\kappa_{i_2\bar{m}}-(A^\kappa_{\bar{m}i_K}-A_{i_2i_K}) \\ \vdots & \ddots & \vdots \\ A^\kappa_{\bar{m}\bar{m}}-A^\kappa_{i_K\bar{m}}-(A^\kappa_{\bar{m}i_1}-A^\kappa_{i_Ki_1}) & \cdots & A^\kappa_{\bar{m}\bar{m}}-A^\kappa_{i_K\bar{m}}-(A^\kappa_{\bar{m}i_K}-A^\kappa_{i_Ki_K}) \end{pmatrix}$$

Proof. We have the following equivalence:

$$\pi_\kappa(\bar{m},r)=\pi_\kappa(i_1,r),\cdots,\pi_\kappa(\bar{m},r)=\pi_\kappa(i_K,r),\ r_{\bar{m}}+\sum_{i=1}^{K}r_{i_l}=1,\ \Sigma_r=\{\bar{m},i_1,\cdots,i_K\}$$

have a unique solution if and only if

$$\pi_\kappa(\bar{m},(1-\sum_{l=1}^K r_{i_l})e_{\bar{m}}+\sum_{l=1}^K r_{i_l}e_{i_l})-\pi_\kappa(i_1,(1-\sum_{l=1}^K r_{i_l})e_{\bar{m}}+\sum_{l=1}^K r_{i_l}e_{i_l})=0,\cdots,$$
$$\pi_\kappa(\bar{m},(1-\sum_{l=1}^K r_{i_l})e_{\bar{m}}+\sum_{l=1}^K r_{i_l}e_{i_l})-\pi_\kappa(i_K,(1-\sum_{l=1}^K r_{i_l})e_{\bar{m}}+\sum_{l=1}^K r_{i_l}e_{i_l})=0$$

have a unique solution. Let fix k. Then we have

$$\begin{aligned}&\pi_\kappa(i_k,(1-\sum_{l=1}^K r_{i_l})e_{\bar{m}}+\sum_{l=1}^K r_{i_l}e_{i_l})-\pi_\kappa(\bar{m},(1-\sum_{l=1}^K r_{i_l})e_{\bar{m}}+\sum_{l=1}^K r_{i_l}e_{i_l})\\=&A^\kappa_{i_k\bar{m}}-A^\kappa_{\bar{m}\bar{m}}+\sum_{l=1}^K((A^\kappa_{\bar{m}\bar{m}}-A^\kappa_{i_k\bar{m}})-(A^\kappa_{\bar{m}i_l}-A^\kappa_{i_ki_l}))r_{i_l}\end{aligned}$$

and from this, we obtain the desired result.

□

Let $\mathcal{K}^*_{\bar{m}}$ be the set of all paths in $\mathcal{K}_{\bar{m}}$ that satisfy the conditions in Proposition B.4. Then, we obviously have

$$\min\{\omega(t):t\in\mathcal{K}^*_{\bar{m}}\}=\min\{\omega(t):t\in\mathcal{K}_{\bar{m}}\}$$

Next, suppose that $\mathbf{q}^*$ is the exit point of the minimum escaping path. If $\pi_\beta(\bar{m},\mathbf{q}^*)=\pi_\beta(i,\mathbf{q}^*)$ for some i, then $\pi_\alpha(\bar{m},\mathbf{q}^*)>\pi_\alpha(l,\mathbf{q}^*)$ for all l and vice versa. This is because if $\pi_\beta(\bar{m},\mathbf{q}^*)=\pi_\beta(i,\mathbf{q}^*)$ and $\pi_\alpha(\bar{m},\mathbf{q}^*)=\pi_\alpha(l,\mathbf{q}^*)$, then we can always construct the escaping path with a smaller cost by removing α-agents' (or β-agents') transitions. Thus, Proposition B.4 implies that if $\pi_\beta(\bar{m},\mathbf{q}^*)=\pi_\beta(i,\mathbf{q}^*)$ for some i, $t_j^{\alpha^*}=0$ for all j.

Proposition B.5 (One-population mistakes)**.** *Suppose that Condition **B** holds. Then there exists t^* such that $\omega(t^*)=\min\{\omega(t):t\in\mathcal{K}^*_{\bar{m}}\}$ and t^* involves only mistakes of one population.*

Proof. Let t^* that satisfies Proposition B.4 be given. Suppose that $t_i^{\alpha^*}>0$. The other case follows similarly. Then, by Proposition B.4, $\pi_\beta(\bar{m},\mathbf{q}^*)=\pi_\beta(i,\mathbf{q}^*)$ for some i. From the remarks before the proposition, we have $\pi_\alpha(\bar{m},\mathbf{q}^*)>\pi_\alpha(l,\mathbf{q}^*)$ for all l. Again, Proposition B.4 implies that $t_l^{\beta^*}=0$ for all l. □

Finally, we have the following result.

Proposition B.6. *Suppose that Condition **B** holds. Then there exists t^* such that* $\min\{\omega(t):\zeta(t)\in\mathcal{K}^*_{\bar{m}}\}$ *and*

$$t_k^{\alpha^*}>0 \textit{ for some } k \textit{ and } t_k^{\alpha^*}=0 \textit{ for all } k\neq l$$

or

$$t_k^{\beta^*}>0 \textit{ for some } k \textit{ and } t_k^{\beta^*}=0 \textit{ for all } k\neq l$$

Proof. Suppose that the minimum cost escaping path involves only one population, say α-population, by Proposition B.5. Then, $x_\beta=e_\beta^{\bar{m}}$ for all x in the minimum cost escaping path. Thus we have $\pi_\alpha(i,x)=$

$\pi_\alpha(j, x)$ for all $i, j \neq \bar{m}$ and for all x in the minimum cost escaping path. The costs of intermediate states in the minimum cost escaping path are the same; the **WBP** implies that the minimum cost escaping path lies in at the boundary of the simplex, yielding the desired result. □

Now the proof for Theorem 5.1 follows from Proposition B.6.

C. Stochastic stability: the maximin criterion

In this section, we examine the problem of finding a stochastically stable state (Foster and Young, 1990). When $\beta = \infty$, the strategy updating dynamic is called an unperturbed process, where each convention becomes an absorbing state for the dynamic. For all $\beta < \infty$, since the dynamic is irreducible, there exists a unique invariant measure. As the noise level becomes negligible ($\beta \to \infty$), the invariant measure converges to a point mass on one of the absorbing states, called a stochastically stable state. One popular way to identify a stochastically stable state is the so-called "maxmin criterion"[7]; when some sufficient conditions are satisfied, this method, along with our results on the exit problem (Theorems 4.1 and 5.1), provides the characterization of stochastic stability.

To study stochastic stability, we have to find a minimum cost path from one convention to another. More precisely, we fix conventions i and j. For one-population models, we let the set of all paths from convention i to j be

$$\mathcal{L}_{i,j}^{(n)} := \{\gamma : \gamma = (x_0, \cdots, x_T) \text{ and } x_0 = e_i,\ x_{t+1} = (x_t)^{k,l}, \text{ for some } k, l, \text{ for all } t < T-1, \\ x_T \in D(e_j) \text{ for some } T > 0\}.$$

We define a similar set for two-population models. We then consider the following problem:

$$C_{ij}^{(n)} := \min\{I^{(n)}(\gamma) : \gamma \in \mathcal{L}_{i,j}^{(n)}\}. \tag{C.1}$$

Again, when n is finite, $C_{ij}^{(n)}$ is complicated, involving many negligible terms; we thus study the stochastic stability problem at $n = \infty$, which again provides the asymptotics of the invariant measure and stochastic stability when n is large. We let

$$C_{ij} = \lim_{n\to\infty} \frac{1}{n} C_{ij}^{(n)} \tag{C.2}$$

and C be a $|S| \times |S|$ matrix whose elements are given by C_{ij} for $i \neq j$ (we set an arbitrary number if $i = j$). Having solved the problems in equation (C.1) (and (C.2)), the standard method to find a stochastically stable state is to construct an $i-$ rooted tree with vertices consisting of the absorbing states and whose cost is defined as the sum of all costs between the absorbing states connected by edges. Then, the stochastic stable state is precisely the root of the minimal cost tree from among all possible rooted trees (see Young (1998b) for more details). In principle, to find a minimal cost tree (hence a stochastically stable state), we need to explicitly solve the problem in equation (C.1). However, in many interesting applications such as bargaining problems, the minimum cost estimates of the escaping path in Theorem 4.1 are sufficient to determine stochastic stability without knowing the true costs of transition between conventions; this method

[7]See Young (1993b, 1998b); Kandori and Rob (1998); Binmore et al. (2003); Hwang et al. (2018)

is called the "maxmin" criterion (see the papers cited in footnote 7; see also Proposition C.1 below). More precisely, we define the incidence matrix of matrix C, $\mathbf{Inc}(C)$, as follows:

$$(\mathbf{Inc}(C))_{ij} := \begin{cases} 1 \text{ if } j = \arg\min_{l\neq i} C_{il} \\ 0 \text{ otherwise} \end{cases}$$

In words, the incidence matrix of C has 1 at the i-th and j-th position if the minimum of elements in the ith row achieves at the i-th and j-th position, and 0 otherwise. We also say that the incidence matrix of C contains a cycle, $(i, i_1, i_2, \cdots, i_{t-1}, i)$, if

$$\mathbf{Inc}(C)_{ii_1}\mathbf{Inc}(C)_{i_1i_2}\cdots\mathbf{Inc}(C)_{i_{t-1}i} > 0$$

for $t \geq 2$. Observe that we can obtain a graph by connecting the vertices of conventions i, j whose $(\mathbf{Inc}(C))_{ij}$ is 1. Also, $\mathbf{Inc}(C)$ always contains a cycle and hence the graph contains the corresponding cycle. If this cycle is unique, by removing an edge from the cycle, we can obtain a tree; this is a candidate tree to the problem of finding a minimal cost tree. Now, we are ready to state some known sufficient conditions to identify stochastic stable states.

Proposition C.1 (Binmore et al. (2003))**.** *Let* $i^* \in \arg\max_i \min_{j\neq i} C_{ij}$. *Suppose that either*
(i) $\max_{j\neq i} C_{ji^*} < \min_{j\neq i} C_{i^*j}$
or
(ii) $\mathbf{Inc}(C)$ *has a unique cycle containing* i^*.
Then i^* *is stochastically stable.*

Proof. See Binmore et al. (2003) □

The sufficient conditions (i) and (ii) for stochastic stability in Proposition C.1 are called the "local resistance test" and "naive minimization test," respectively (Binmore et al., 2003). If strategy i pairwisely risk-dominates strategy j (i.e., $A_{ii} - A_{ji} > A_{jj} - A_{ji}$), then under the uniform mistake model, $C_{ij} > 1/2$ and $C_{ji} < 1/2$ hold. Thus, if strategy i^* pairwisely risk-dominates all strategies (called a globally pairwise risk-dominant strategy), then $C_{i^*j} > 1/2$ for all $j \neq i$ and $C_{ji^*} < 1/2$ for all $j \neq i$. Thus condition (i) in Proposition C.1 holds and i^* is stochastically stable (see Theorem 1 in Kandori and Rob (1998) and Corollary 1 in Ellison (2000)).

The number $\min_{j\neq i} C_{ij}$ in Proposition C.1 is, as mentioned, often called the "radius" of convention i; this measures how difficult it is to escape from convention i (Ellison, 2000). Proposition C.1 shows that if either (i) or (ii) holds, the state with the greatest radius (and hence the state most difficult to escape) is stochastically stable. To check whether either condition (i) or (ii) holds, clearly it is enough to know that $\min_{j\neq i} C_{ij}, \max_{j\neq i} C_{ji}$ etc.

An important consequence of our main theorem on the exit problem (Theorem 4.1) is that it provides the lower and upper bounds of the radius of convention i, $\min_{j\neq i} C_{ij}$, as follows. On the one hand, a path escaping from convention i to j (in $\mathcal{L}^{(n)}_{i,j}$) by definition exits the basin of attraction of convention i and thus $\mathcal{L}^{(n)}_{i,j} \subset \mathcal{G}^{(n)}_i$ in equation (30). Thus,

$$C^{(n)}_{ij} = \min\{I^{(n)}(\gamma) : \gamma \in \mathcal{L}^{(n)}_{i,j}\} \geq \min\{I^{(n)}(\gamma) : \gamma \in \mathcal{G}^{(n)}_i\}, \tag{C.3}$$

and Theorem 4.1 shows that

$$\lim_{n\to\infty}\frac{1}{n}\min\{I^{(n)}(\gamma):\gamma\in\mathcal{G}_i^{(n)}\}=\min_{j\neq i}R_{ij}. \tag{C.4}$$

Then equations (C.3) and (C.4) together give a lower bound for $\min_{j\neq i} C_{ij}$. On the other hand, if $\gamma_{i\to j}$ is the straight line path from convention i to j ending at the mixed strategy Nash equilibrium involving i and j, we have

$$I^{(n)}(\gamma_{i\to j})\geq\min\{I^{(n)}(\gamma):\gamma\in\mathcal{L}_{i,j}^{(n)}\}=C_{ij}^{(n)} \tag{C.5}$$

and

$$\lim_{n\to\infty}\frac{1}{n}I^{(n)}(\gamma_{i\to j})=R_{ij}. \tag{C.6}$$

Thus, equations (C.5) and (C.6) give an upper bound for $\min_{j\neq i} C_{ij}$. These are the main contents of the following proposition.

Proposition C.2. *Suppose Condition **A** or Condition **B** holds. Then*
(i) $C_{ij}\leq R_{ij}$ *for all* i,j.
(ii) $\min_{j\neq i} C_{ij}=\min_{j\neq i} R_{ij}$.
(iii) $\arg\min_{j\neq i} R_{ij}\subset\arg\min_{j\neq i} C_{ij}$ *for all* i.

Proof. We obtain (i) by dividing equation (C.5) by n , taking the limit, and using (C.6). For (ii), from equations (C.3) and (C.4), $\lim_{n\to\infty}\frac{1}{n}C_{ij}^{(n)}\geq\min_{j\neq i} R_{ij}$, implying that $\min_{j\neq i} C_{ij}\geq\min_{j\neq i} R_{ij}$. Also from (i), we have $\min_{j\neq i} C_{ij}\leq\min_{j\neq i} R_{ij}$. Thus, (ii) follows. We next prove (iii). Suppose that $j^{**}\in\arg\min_{j\neq i} R_{ij}$ and $j^*\in\arg\min_{j\neq i} C_{ij}$. Then from (i) and (ii), $R_{ij^{**}}=C_{ij^*}\leq C_{ij^{**}}\leq R_{ij^{**}}$. Thus $j^{**}\in\arg\min_{j\neq i} C_{ij}$ and we have $\arg\min_{j\neq i} R_{ij}\subset\arg\min_{j\neq i} C_{ij}$. □

The immediate consequence of Proposition C.2 is that $\arg\max_i\min_{j\neq i} C_{ij}=\arg\max_i\min_{j\neq i} R_{ij}$ and $\max_{j\neq i} C_{ji}\leq\max_{j\neq i} R_{ji}$. Further, if $\arg\min_{j\neq i} C_{ij}$ is unique for all i, from Proposition C.2, the incidence matrices of C and R are the same. In general, $\arg\min_{j\neq i} C_{ij}$ may not be unique for some i. In this case, Proposition C.2 (iii) implies that if $R_{ij}=1$, then $C_{ij}=1$, which, in turn, implies that whenever R yields a graph containing a unique cycle, C yields the same graph containing the unique cycle. These facts enable us to replace C in Proposition C.1 by R—a $|S|\times|S|$ matrix consisting of R_{ij}s (again, we assign arbitrary numbers at the diagonal positions). This is our main result on stochastic stability.

Theorem C.1 (**Stochastic Stability**). *Suppose that Condition **A** or Condition **B** holds. Let* $i^*\in\arg\max_i\min_{j\neq i} R_{ij}$. *Suppose also that either*
(i) $\max_{j\neq i} R_{ji^*}<\min_{j\neq i} R_{i^*j}$
or
(ii) $\mathbf{Inc}(R)$ *has a unique cycle containing* i^*.
Then, i^* *is stochastically stable.*

Proof. Let $i^*\in\arg\max_i\min_{j\neq i} R_{ij}$. From Proposition C.2 (iii), $i^*\in\arg\max_i\min_{j\neq i} C_{ij}$. We first suppose that (i) holds. Now, Propositions C.2 (i) and C.2 (ii) imply that

$$\max_{j\neq i^*} C_{ji^*}\leq\max_{j\neq i^*} R_{ji^*}<\min_{j\neq i^*} R_{i^*j}=\min_{j\neq i^*} C_{i^*j}.$$

Thus, Proposition C.1 implies that i^* is stochastically stable. Now, suppose that (ii) holds. From Proposition C.2 (iii) and the remarks before Theorem C.1, **Inc**(C) contains a unique cycle containing i^*, too. Thus, Proposition C.1 again implies that i^* is stochastically stable. □

Note that two-strategy games trivially satisfy both conditions (i) and (ii) in Theorem C.1. Here, we can easily check that the stochastic stable state is the risk-dominant equilibrium. In particular, Kandori and Rob (1998) show that when a coordination game exhibits positive feedback (the marginal bandwagon property), a "globally pairwise risk-dominant equilibrium" is stochastically stable under the uniform mistake model (see also Binmore et al. (2003)). However, when the number of strategies exceeds two, Theorem C.1 shows that stochastically stable states under the logit choice rule do not necessary satisfy the criterion of pairwise risk dominance. To summarize, Theorem C.1 asserts that when either condition (i) or condition (ii) is satisfied, the state with the largest radius (and hence the most difficult state to escape) is stochastically stable, in line with the existing results for uniform interaction models. However, the radius now depends on the opportunity cost of individuals' mistakes as well as the threshold number of agents inducing others to play a new best-response.

D. Stochastic stable states for Nash demand games

We first show that Nash demand game,

$$(A^{\alpha}_{ij}, A^{\beta}_{ij}) := \begin{cases} (\delta i, f(\delta j)), & \text{if } i \leq j \\ (0,0), & \text{if } i > j, \end{cases} \tag{D.1}$$

satisfies **Condition B**.

Condition B (i).
We divide cases as follows:
(1) $\bar{m} > i > j$.

$$A^{\alpha}_{\bar{m}\bar{m}} - A^{\alpha}_{i\bar{m}} - (A^{\alpha}_{\bar{m}j} - A^{\alpha}_{ij}) = \delta m - \delta i > 0,\ \ A^{\beta}_{\bar{m}\bar{m}} - A^{\beta}_{\bar{m}i} - (A^{\beta}_{j\bar{m}} - A^{\beta}_{ij}) = f(\delta\bar{m}) - (f(\delta\bar{m}) - f(\delta i)) > 0$$

(2) $\bar{m} > j > i$.

$$A^{\alpha}_{\bar{m}\bar{m}} - A^{\alpha}_{i\bar{m}} - (A^{\alpha}_{\bar{m}j} - A^{\alpha}_{ij}) = \delta m - \delta i + \delta_j > 0,\ \ A^{\beta}_{\bar{m}\bar{m}} - A^{\beta}_{\bar{m}i} - (A^{\beta}_{j\bar{m}} - A^{\beta}_{ij}) = f(\delta\bar{m}) - f(\delta\bar{m}) > 0$$

(3) $i > \bar{m} > j$.

$$A^{\alpha}_{\bar{m}\bar{m}} - A^{\alpha}_{i\bar{m}} - (A^{\alpha}_{\bar{m}j} - A^{\alpha}_{ij}) = \delta m > 0,\ \ A^{\beta}_{\bar{m}\bar{m}} - A^{\beta}_{\bar{m}i} - (A^{\beta}_{j\bar{m}} - A^{\beta}_{ij}) = f(\delta\bar{m}) - f(\delta i) - (f(\delta\bar{m}) - f(\delta i)) = 0$$

(4) $j > \bar{m} > i$.

$$A^{\alpha}_{\bar{m}\bar{m}} - A^{\alpha}_{i\bar{m}} - (A^{\alpha}_{\bar{m}j} - A^{\alpha}_{ij}) = \delta\bar{m} - \delta i - (\delta\bar{m} - \delta i) = 0,\ \ A^{\beta}_{\bar{m}\bar{m}} - A^{\beta}_{\bar{m}i} - (A^{\beta}_{j\bar{m}} - A^{\beta}_{ij}) = f(\delta\bar{m}) > 0$$

(5) $i > j > \bar{m}$.

$$A^{\alpha}_{\bar{m}\bar{m}} - A^{\alpha}_{i\bar{m}} - (A^{\alpha}_{\bar{m}j} - A^{\alpha}_{ij}) = \delta\bar{m} - \delta\bar{m} = 0,\ \ A^{\beta}_{\bar{m}\bar{m}} - A^{\beta}_{\bar{m}i} - (A^{\beta}_{j\bar{m}} - A^{\beta}_{ij}) = f(\delta\bar{m}) - f(\delta i) - (-f(\delta i)) > 0$$

(6) $j > i > \bar{m}$.

$$A^{\alpha}_{\bar{m}\bar{m}} - A^{\alpha}_{i\bar{m}} - (A^{\alpha}_{\bar{m}j} - A^{\alpha}_{ij}) = \delta\bar{m} - (\delta\bar{m} - \delta i) > 0,\ \ A^{\beta}_{\bar{m}\bar{m}} - A^{\beta}_{\bar{m}i} - (A^{\beta}_{j\bar{m}} - A^{\beta}_{ij}) = f(\delta\bar{m}) - f(\delta i) > 0$$

Condition B (ii).
We first show the following lemma.

Lemma D.1. *Suppose that A is a $n \times n$ matrix such that*

$$A_{ij} = a_i \text{ if } i \leq j,\ = 0 \text{ if } i > j,\ a_i < a_{i+1} \text{ for all } i = 1, \cdots, n-1$$

Then there exists a unique $x \gg 0$ such that $Ax = \mathbf{1}$ where $\mathbf{1}$ is the column vector consisting all 1's.

Proof. Let x be
$$x^T = \left(\frac{1}{a_1} - \frac{1}{a_2}, \cdots, \frac{1}{a_{n-1}} - \frac{1}{a_n}, \frac{1}{a_n}\right)$$
Note that by the assumption, we have $x \gg 0$. Then we have

$$(Ax)_k = \sum_{i=1}^{n} A_{ki}x_i = \sum_{i=1}^{k} a_k x_i = a_k \sum_{i=k}^{n} x_i = a_k \frac{1}{a_k} = 1$$

Suppose that there exists y such that $Ay = \mathbf{1}$. Then, since $det(A) \neq 0$, $y = A^{-1}\mathbf{1} = x$. Thus $x \gg 0$ is unique. □

Now let $i_1, \cdots, i_K$. We rearrange i_k's such that $i_1 < \cdots < i_K$. Let A be a matrix whose rows and columns consist of $i_1, \cdots, i_K$. Then from (D.1), the hypothesis of Lemma D.1 is satisfied. Thus, by normalizing x, we can find a unique $q \in \Delta_\beta$ which satisfies the desired property.

Recall that

$$R^U_{mj} := \min\{(A^{\beta}_{mm} - A^{\beta}_{mj})\frac{(A^{\alpha}_{mm} - A^{\alpha}_{jm})}{(A^{\alpha}_{mm} - A^{\alpha}_{jm}) + (A^{\alpha}_{jj} - A^{\alpha}_{mj})}, (A^{\alpha}_{mm} - A^{\alpha}_{jm})\frac{(A^{\beta}_{mm} - A^{\beta}_{mj})}{(A^{\beta}_{mm} - A^{\beta}_{mj}) + (A^{\beta}_{jj} - A^{\beta}_{jm})}\}$$

and
$$(A^{\alpha}_{ij}, A^{\beta}_{ij}) := \begin{cases} (\delta i, f(\delta j)), & \text{if } i \leq j \\ (0, 0), & \text{if } i > j \end{cases}$$

Then we divide cases:

(i) $m < j$. We find that

$$A^{\beta}_{mm} = f(\delta m), A^{\beta}_{mj} = f(\delta j), A^{\alpha}_{mm} = \delta m, A^{\alpha}_{jm} = 0, A^{\alpha}_{jj} = \delta j, A^{\alpha}_{mj} = \delta m$$

and

$$A^{\alpha}_{mm} = \delta m, A^{\alpha}_{jm} = 0, A^{\beta}_{mm} = f(\delta m), A^{\beta}_{mj} = f(\delta j), A^{\beta}_{jm} = 0, A^{\beta}_{jj} = f(\delta j)$$

Using these, we find that

$$(A^{\beta}_{mm} - A^{\beta}_{mj})\frac{A^{\alpha}_{mm} - A^{\alpha}_{jm}}{A^{\alpha}_{mm} - A^{\alpha}_{jm} + (A^{\alpha}_{jj} - A^{\alpha}_{mj})} = (f(\delta m) - f(\delta j))\frac{\delta m}{\delta j}$$

and

$$(A^{\alpha}_{mm} - A^{\alpha}_{jm})\frac{A^{\beta}_{mm} - A^{\beta}_{mj}}{A^{\beta}_{mm} - A^{\beta}_{mj} + (A^{\beta}_{jj} - A^{\beta}_{jm})} = \delta m \frac{f(\delta m) - f(\delta j)}{f(\delta m)}.$$

(ii) $m > j$. We find that

$$A^{\beta}_{mm} = f(\delta m), A^{\beta}_{mj} = 0, A^{\alpha}_{mm} = \delta m, A^{\alpha}_{jm} = \delta j, A^{\alpha}_{jj} = \delta j, A^{\alpha}_{mj} = 0$$

and

$$A^{\alpha}_{mm} = \delta m, A^{\alpha}_{jm} = \delta j, A^{\beta}_{mm} = f(\delta m), A^{\beta}_{mj} = 0, A^{\beta}_{jm} = f(\delta m), A^{\beta}_{jj} = f(\delta j)$$

Using these, we find that

$$(A^{\beta}_{mm} - A^{\beta}_{mj})\frac{A^{\alpha}_{mm} - A^{\alpha}_{jm}}{A^{\alpha}_{mm} - A^{\alpha}_{jm} + (A^{\alpha}_{jj} - A^{\alpha}_{mj})} = f(\delta m)\frac{\delta m - \delta j}{\delta m}$$

and

$$(A^{\alpha}_{mm} - A^{\alpha}_{jm})\frac{A^{\beta}_{mm} - A^{\beta}_{mj}}{A^{\beta}_{mm} - A^{\beta}_{mj} + (A^{\beta}_{jj} - A^{\beta}_{mj})} = (\delta m - \delta j)\frac{f(\delta m)}{f(\delta j)}.$$

Thus we have

$$R^{U}_{mj} = \begin{cases} (f(\delta m) - f(\delta j))\frac{\delta m}{\delta j} \wedge \delta m \frac{f(\delta m) - f(\delta j)}{f(\delta m)} & \text{if } m < j \\ f(\delta m)\frac{\delta m - \delta j}{\delta m} \wedge (\delta m - \delta j)\frac{f(\delta m)}{f(\delta j)} & \text{if } m > j \end{cases}$$

Or

$$R^{U}_{mj} = \min_{m<j}\{(f(\delta m) - f(\delta j))\frac{\delta m}{\delta j} \wedge \delta m \frac{f(\delta m) - f(\delta j)}{f(\delta m)}\} \wedge \min_{m>j}\{f(\delta m)\frac{\delta m - \delta j}{\delta m} \wedge (\delta m - \delta j)\frac{f(\delta m)}{f(\delta j)}\}.$$

Note that we have

$$R^{I}_{mj} = \min_{m<j}\{\delta m \frac{f(\delta m) - f(\delta j)}{f(\delta m)}\} \wedge \min_{m>j}\{f(\delta m)\frac{\delta m - \delta j}{\delta m}\}.$$

Then we would like to find $\min_j R^{U}_{mj}$. To do this, we first have the following lemma.

Lemma D.2. *Suppose that $f(x) \geq 0$, $f'(x) < 0$ and $f''(x) < 0$ for all x. Let y be given.*
(i) $\frac{f'(x)}{f(x)}$ is decreasing in x.
(ii) $xf'(x) - f(x)$ is decreasing in x.
(iii) $f'(x) + \frac{f(x)}{x}$ is decreasing in x.
(iv) $f'(x) + (\frac{f(x)}{x})^2$ is decreasing in x.
(v) $(f(y) - f(x))\frac{y}{x}$ is increasing in x
(vi) $(y - x)\frac{f(y)}{f(x)}$ is decreasing in x.

Proof. (i)-(iv) are easily verified by taking derivatives. We show (v). (vi) follows similarly. Let $\varphi(x) := (f(y) - f(x))\frac{y}{x}$. We find that

$$\varphi'(x) = y\frac{-f'(x)x + f(x) - f(y)}{x^2}$$

Then since $-f'(x)x + f(x)$ is increasing in x, we have

$$-f'(x)x + f(x) - f(y) \geq f(0) - f(y) \geq 0$$

since f is decreasing. Thus $\varphi'(x) > 0$. □

Thus using Lemma (D.2), we find that

$$\min_j R^U_{mj} = \min\{(f(\delta m) - f(\delta(m+1)))\frac{\delta m}{\delta(m+1)}, \delta m\frac{f(\delta m) - f(\delta(m+1))}{f(\delta(m))}, f(\delta m)\frac{\delta}{\delta m}, \delta\frac{f(\delta m)}{f(\delta(m-1))}\}$$

We let

$$r_1(m) := (f(\delta m) - f(\delta(m+1)))\frac{\delta m}{\delta(m+1)},\ r_2(m) := \delta m\frac{f(\delta m) - f(\delta(m+1))}{f(\delta(m))}$$

and

$$l_1(m) := f(\delta m)\frac{\delta}{\delta m},\ l_2(m) := \delta\frac{f(\delta m)}{f(\delta(m-1))}.$$

Lemma D.3. *We have the following results:*
(i) r_1 and r_2 are increasing in m.
(ii) l_1 and l_2 are decreasing in m.

Proof. (i). Since $f'' < 0$, $f(\delta m) - f(\delta(m+1))$ is increasing. Since $\frac{\delta m}{\delta(m+1)}$ is increasing, two terms in r_1 are both positive and increasing, hence r_1 is increasing. Also since $f'' < 0$, $\frac{f(\delta(m+1))}{f(\delta m)}$ is decreasing in m. Thus r_2 is increasing. □

Then r_1 and r_2 are increasing in m and l_1 and l_2 are decreasing in m.

Lemma D.4. *Suppose that*

$$m^* \in \arg\max_m \min_j R^U_{mj}$$

Then for all $m < m^*$, $\min_j R^U_{mj} = R^U_{m,m+1}$ and for all $m > m^*$, $\min_j R^U_{mj} = R^U_{m,m-1}$

Proof. Let $\hat{R}(m) := \min_j R^U_{mj}$. We show that

$$\begin{aligned}&\text{If } m < m^*,\ \text{then } \hat{R}(m) = r_1(m)\ \text{ or } r_2(m)\\&\text{If } m > m^*,\ \text{then } \hat{R}(m) = l_1(m)\ \text{ or } l_2(m)\end{aligned}$$

and then the desired results follow. We show the first claim. (the second claim follows similarly). Let $m < m^*$ and $\hat{R}(m) = l_1(m)$. Then since $l_1(m)$ is decreasing in m, $l_1(m) > l_1(m^*)$ and by definition, we have $\hat{R}(m^*) \leq l_1(m^*)$. Thus we have

$$\min_j R^U_{mj} = \hat{R}(m) = l_1(m) > l_1(m^*) \geq \hat{R}(m^*)$$

which is contradiction to $m^* \in \arg\max_m \min_j R^U_{mj}$ If $\hat{R}(m) = l_2(m)$, the exactly same argument leads to a contradiction. Thus if $m < m^*$, then $\hat{R}(m) = r_1(m)$ or $r_2(m)$. □

Let s^* and s^I such that

$$-f'(s^*) = \frac{f(s^*)}{s^*} \text{ and } -f'(s^I) = (\frac{f(s^I)}{s^I})^2$$

and for $\mu \in [0, \frac{\bar{s}_\alpha}{\delta}] \cap \mathbb{R}$, let $\mu^I = \mu^I(\delta)$, $\mu^* = \mu^*(\delta)$, and $\mu^{**} = \mu^{**}(\delta)$ such that

$$r_1(\mu^*) = l_1(\mu^*), r_2(\mu^{**}) = l_2(\mu^{**}) \text{ and } r_2(\mu^I) = l_1(\mu^I). \tag{D.2}$$

Lemma D.5. *We have the following results. As $\delta \to 0$,*

$$\delta\mu^*(\delta) \to s^*, \qquad \delta\mu^{**}(\delta) \to s^*, \qquad \delta\mu^I(\delta) \to s^I.$$

Proof. For $\delta\mu^*(\delta) \to s^*$, let

$$\varphi_\delta(x) := \frac{(f(x) - f(x+\delta))}{\delta}\frac{x^2}{(x+\delta)f(x)}, \quad \varphi(x) := -f'(x)\frac{x}{f(x)}.$$

Then φ_δ converge uniformly to φ and $\varphi_\delta(\delta\mu^*(\delta)) = \frac{r_1(\mu^*)}{l_1(\mu^*)} = 1$ and $\varphi(x^*) = 1$. Then the uniform convergence of φ_δ to φ implies that $\delta\mu^*(\delta) \to s^*$. The second and third parts follow similarly. □

Next we show that

Lemma D.6. *We have the following result.*
(i) If $s^* > s^E$, then $s^* > s^I > s^E$ and $-f'(s^I)\frac{s^I}{f(s^I)} < 1$ and $-f'(s^*) < 1$
(ii) If $s^* < s^E$, then $s^* < s^I < s^E$ and $-f'(s^I)\frac{s^I}{f(s^I)} > 1$ $-f'(s^*) > 1$

Proof. We show (i) and (ii) follows similarly. Suppose that $s^* > s^E$. Let $s^I \geq s^*$. Since from Lemma D.2 $-f'(x) - \frac{f(x)}{x}$ is increasing, we have

$$-f'(s^I) - \frac{f(s^I)}{s^I} \geq -f'(s^*) - \frac{f(s^*)}{s^*} = 0 = -f'(s^I) - (\frac{f(s^I)}{s^I})^2$$

which implies that

$$\frac{f(s^I)}{s^I} \geq 1 = \frac{f(s^E)}{s^E}$$

Since $\frac{f(s)}{s}$ is decreasing in s, we have

$$s^E \geq s^I \geq s^* > s^E$$

which is a contradiction. Now suppose that $s^I \leq s^E$. Then since $s^E < s^*$,

$$-f'(s^I) - \frac{f(s^I)}{s^I} < -f'(s^*) - \frac{f(s^*)}{s^*} = 0 = -f'(s^I) - (\frac{f(s^I)}{s^I})^2$$

which implies that

$$\frac{f(s^I)}{s^I} < 1.$$

which is a contradiction to $\frac{f(s^I)}{s^I} \geq \frac{f(s^E)}{s^E} = 1$ from $s^I \leq s^E$. Now from $s^* > s^I$ and $s^* > s^E$, respectively we have

$$-f'(s^I)\frac{s^I}{f(s^I)} < 1 \text{ and } -f'(s^*) < 1.$$

□

Lemma D.7. *We have the following results.*
(i) If $s^* > s^E$, then there exists $\underline{\delta}$ such that for all $\delta < \underline{\delta}$, $\mu^* > \mu^I$ and

$$r_1(\mu^I) < r_2(\mu^I) = l_1(\mu^I) \text{ and } r_1(\mu^*) < l_2(\mu^*)$$

where $\mu^I = \mu^I(\delta)$ and $\mu^* = \mu^*(\delta)$ are defined in (D.2).
(ii) If $s^* < s^E$, then there exists $\underline{\delta}$ such that for all $\delta < \underline{\delta}$, $\mu^{**} < \mu^I$ and

$$l_2(\mu^I) < r_2(\mu^I) = l_1(\mu^I) \text{ and } l_2(\mu^{**}) < r_1(\mu^{**})$$

where $\mu^I = \mu^I(\delta)$ and $\mu^* = \mu^*(\delta)$ are defined in (D.2).

Proof. We first prove (i). Suppose that $s^* > s^E$. From Lemma D.6, we have

$$-f'(s^I)\frac{s^I}{f(s^I)} < 1 \text{ and } -f'(s^*) < 1 \tag{D.3}$$

Since $\delta\mu^I \to s^I$ (Lemma D.5) and $s^I < s^*$ and from (D.3)

$$\frac{r_1(\mu^I)}{l_1(\mu^I)} \to -f'(s^I)\frac{s^I}{f(s^I)} < 1,$$

there exists $\underline{\delta}$ such that for all $\delta < \underline{\delta}$, $r_1(\mu^I) < l_1(\mu^I)$ and $\mu^I < \mu^*$. For the second inequality $r_1(\mu^*) < l_2(\mu^*)$ similarly follows from

$$\frac{r_1(\mu^*)}{l_2(\mu^*)} < 1 \iff \frac{f(\delta\mu^*) - f(\delta(\mu^*+1))}{\delta}\frac{\delta\mu^*}{\delta(\mu^*+1)}\frac{f(\delta(\mu^*-1))}{f(\delta\mu^*)} < 1$$

and

$$\frac{r_1(\mu^*)}{l_2(\mu^*)} \to -f'(s^*) < 1$$

from (D.3).

Next we show (ii). Similarly to (i), from Lemma D.6, we have we have

$$-f'(s^I)\frac{s^I}{f(s^I)} > 1 \text{ and } f'(s^*) > 1$$

Then we have

$$\frac{l_2(\mu^I)}{r_2(\mu^I)} < 1 \iff \frac{\delta}{f(\delta\mu^I) - f(\delta(\mu^I+1))}\frac{f(\delta\mu^I)}{f(\delta(\mu^I-1))}\frac{f(\delta\mu)}{\delta\mu} < 1$$

and

$$\frac{l_2(\mu^{**})}{r_1(\mu^{**})} < 1 \iff \frac{\delta}{f(\delta\mu^{**}) - f(\delta(\mu^{**}+1))}\frac{\delta(\mu^{**}+1)}{\delta\mu^{**}}\frac{f(\delta\mu^{**})}{f(\delta(\mu^{**}-1))} < 1$$

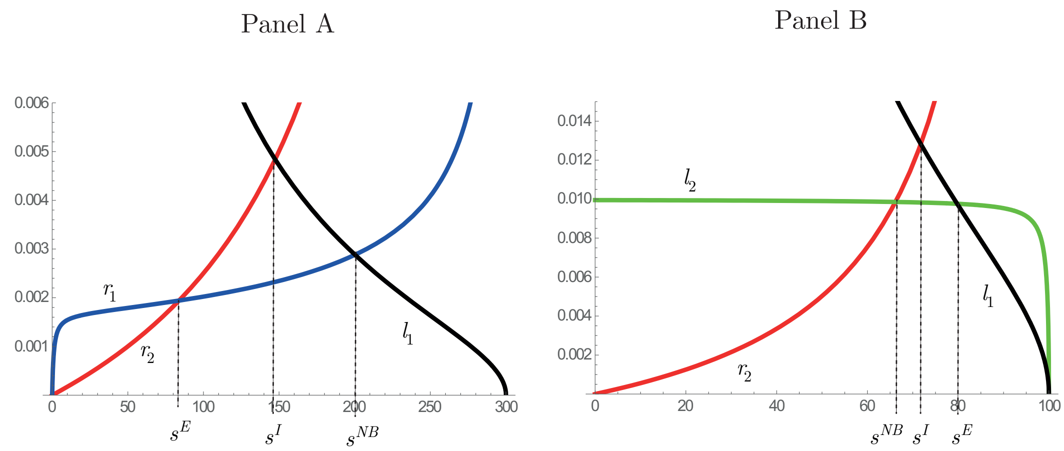

Figure D.8: **Determinations of stochastically stable states.** For Panel A, $f(x) = \sqrt{1 - \frac{x}{3}}$ for $x \in [0, 3]$, $\delta = 0.01$. For Panel B, $f(x) = \sqrt{3(1-x)}$, for $x \in [0, 1]$, $\delta = 0.01$.

and from these, (ii) follows. $\square$

Lemma D.8. *Suppose that* μ^* *is given by* (D.2).
(i) If $s^* > s^E$*, then*

$$\mu^* \in \arg \max_{\mu \in [0, \frac{\bar{s}}{\delta}]} \min\{r_1(\mu), r_2(\mu), l_1(\mu), l_2(\mu)\}$$

(ii) If $s^* < s^E$*, then*

$$\mu^{**} \in \arg \max_{\mu \in [0, \frac{\bar{s}}{\delta}]} \min\{r_1(\mu), r_2(\mu), l_1(\mu), l_2(\mu)\}$$

Proof. Let $s^* > s^E$. Choose $\underline{\delta}$ satisfying Lemma D.7. Then for all $\delta < \underline{\delta}$, we have

$$r_1(\mu^*) = l_1(\mu^*) < l_1(\mu_0) = r_2(\mu_0) < r_2(\mu^*)$$

and thus $r_1(\mu^*) \leq \min\{r_2(\mu^*), l_1(\mu^*), l_2(\mu^*)\}$. Now, if $\mu < \mu^*$ then $r_1(\mu^*) > r_1(\mu)$ since $r_1(\cdot)$ is increasing. If $\mu > \mu^*$, then $r_1(\mu^*) = l_1(\mu^*) > l_1(\mu)$ since $l_1(\cdot)$ is decreasing. Thus we have

$$r_1(\mu^*) \geq \min\{r_1(\mu), r_2(\mu), l_1(\mu), l_2(\mu)\}$$

for all $\mu \in [0, \frac{\bar{s}}{\delta}]$. This shows that

$$\mu^* \in \arg \max_{\mu \in [0, \frac{\bar{s}}{\delta}]} \min\{r_1(\mu), r_2(\mu), l_1(\mu), l_2(\mu)\}$$

Now let $s^* < s^E$. Again choose $\underline{\delta}$ satisfying Lemma D.7. Then for all $\delta < \underline{\delta}$, we have

$$r_2(\mu^{**}) = l_2(\mu^{**}) < r_1(\mu^{**}) < r_1(\mu^I) = l_1(\mu^I) < l_1(\mu^{**})$$

and similarly since r_2 is increasing and l_2 is decreasing, we obtain the desired result. $\square$

Thus we have the following result.

	α favored transition		β favored transition		Stochastic stability
	β mistake (A)	α mistake (B)	β mistake (C)	α mistake (D)	
Uniform	$\frac{\delta m}{\bar{s}-\alpha}$	$\frac{\Delta f(\delta m)}{f(\delta m)}$	$\frac{\delta}{\delta m}$	$\frac{f(\delta m)}{f(\delta)}$	
Unintentional Intentional		○	○		$\frac{\Delta f(\delta m)}{f(\delta m)} \approx \frac{\delta}{\delta m}$
Logit Unintentional	$\Delta f(\delta m)\frac{\delta m}{\delta(m+1)}$	$\delta m \frac{\Delta f(\delta m)}{f(\delta(m))}$	$f(\delta m)\frac{\delta}{\delta m}$	$\delta \frac{f(\delta m)}{f(\delta(m-1))}$	
$s^{NB} > s^E$	○	△	○		$\Delta f(\delta m)\frac{\delta m}{\delta(m+1)} \approx f(\delta m)\frac{\delta}{\delta m}$
$s^{NB} < s^E$		○	△	○	$\delta m \frac{\Delta f(\delta m)}{f(\delta(m))} \approx \delta \frac{f(\delta m)}{f(\delta(m-1))}$
Logit Intentional		○	○		$\delta m \frac{\Delta f(\delta m)}{f(\delta(m))} \approx f(\delta m)\frac{\delta}{\delta m}$

Table D.2: Comparison of solutions under various mistake models. $\Delta f(\delta m) := f(\delta m) - f(\delta(m+1))$. Resistances are determined by the minimum of A, B, C, and D. In the rows tilted with "unintentional", "intentional", $s^{NB} > s^E$, $s^{NB} < s^E$, and "logit intentional" show the smaller ones. Thus under the logit unintentional dynamic, when $s^{NB} > s^E$, the transition always occurs by β population, while $s^{NB} < s^E$, the transition always occurs by α population. Entries marked by △ and ○ occurs in the minimal tree, but entries marked by ○ are only binding and hence determining the stochastic stable convention.

Theorem D.1. *Consider the logit choice rule. There exists $\underline{\delta}$ such that for all $\delta < \underline{\delta}$, the stochastic stable state $m^{st}(\delta)$ converges to s^{NB}: i.e.,*

$$\delta m^{st}(\delta) \to s^{NB}$$

where

$$-f'(s^{NB}) = \frac{f(s^{NB})}{s^{NB}}.$$

Proof. Choose $\underline{\delta}$ satisfying Lemma D.7. Let $\delta < \underline{\delta}$. If $s^* > s^E$, then pick $m^{st}(\delta)$ to be the integer closest to $\mu^*(\delta)$ in (D.2). If $s^* < s^E$, the pick $m^{st}(\delta)$ to be the integer closest to $\mu^{**}(\delta)$. Then Lemma D.4, Lemma D.8 and Theorem C.1 show that $m^{st}(\delta)$ is a stochastically stable state. Since $\mu^*(\delta), \mu^*(\delta) \to s^*$, we have $\delta m^{st}(\delta) \to s^* = s^{NB}$ and obtain the desired result. □

Theorem D.2. *Consider the intentional logit choice rule. There exists $\underline{\delta}$ such that for all $\delta < \underline{\delta}$, the stochastic stable state $m^{st}(\delta)$ converges to s^I: i.e.,*

$$\delta m^{st}(\delta) \to s^I$$

where

$$-f'(s^I) = (\frac{f(s^I)}{s^I})^2.$$

Proof. Under the intentional logit choice rule, we have

$$\min_j R^I_{mj} = \min\{\delta m \frac{f(\delta m) - f(\delta(m+1))}{f(\delta m)}, f(m\delta)\frac{\delta}{m\delta}\}$$

Then the exactly same argument as for the unintentional logit choice rule shows the desired result. □