Online Appendix for “Record-Keeping and Cooperation in Large Societies”

Daniel Clark, Drew Fudenberg, and Alexander Wolitzky

January 4, 2021

OA.1 Proof of Corollary 3

Corollary 3. *Under any finite-partitional record system, a coordination-proof equilibrium exists if the stage game has a symmetric Nash equilibrium that is not Pareto-dominated by another (possibly asymmetric) Nash equilibrium.*

Fix such a symmetric static equilibrium α^*, and let σ recommend α^* at every record pair (r, r'). Then (σ, μ) is an equilibrium for any steady state μ. Moreover, note that $\hat{u}_{r,r'}(a, a') = (1-\gamma)u(a, a') + \gamma u(\alpha^*, \alpha^*)$, for any r, r', a, a'. Thus, (α, α') is a (possibly mixed) augmented-game Nash equilibrium if and only if it is a Nash equilibrium of the stage game. Since (α^*, α^*) is not Pareto-dominated by another static equilibrium, there is no augmented-game Nash equilibrium (α, α') satisfying $(u(\alpha, \alpha'), u(\alpha', \alpha)) > (u(\alpha^*, \alpha^*), u(\alpha^*, \alpha^*))$, and hence there is no augmented-game Nash equilibrium (α, α') satisfying $(\hat{u}_{r,r'}(\alpha, \alpha'), \hat{u}_{r',r}(\alpha', \alpha)) > (\hat{u}_{r,r'}(\alpha^*, \alpha^*), \hat{u}_{r',r}(\alpha^*, \alpha^*))$ for any r, r'. That is, (σ, μ) is coordination-proof.

OA.2 Proof of Theorem 3

Theorem 3. *Fix an action a. With canonical first-order records:*

(i) If there exists an unprofitable punishment b for a and there is a strict and symmetric static equilibrium (d, d), then a can be limit-supported by strict equilibria.

(ii) If there exists an action b such that (b, b) is a strict static equilibrium and $u(a, a) >$ $\max\{u(b, a), u(b, b)\}$*, then a can be limit-supported by strict equilibria.*

Let $0 < \underline{\gamma} < \overline{\gamma} < 1$ be such that

$$\frac{\gamma}{1-\gamma} > \max\left\{\max_x \frac{u(x,a)-u(a,a)}{u(a,a)-u(c,b)}, \max_x \frac{u(x,c)-u(b,c)}{u(a,a)-u(c,b)}\right\} \tag{OA 1}$$

for all $\gamma \in [\underline{\gamma}, \overline{\gamma}]$. Consider the strategy $\hat{\sigma}$: A player whose action has never been recorded as anything other than a or b is in good standing, and all other players are in bad standing. Players in good standing play a against fellow good-standing players and play b against bad-standing players, while bad-standing players always play b. described in Section 4, and let μ^G denote the share of good-standing players in a steady state. We will show that for all $\delta > 0$, there is an $\overline{\varepsilon} > 0$ such that, whenever $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ and $\varepsilon_{x,x'} < \overline{\varepsilon}$ for all $x, x' \in A$, $\hat{\sigma}$ induces strict equilibria satisfying $\mu^G > 1-\delta$. Thus, $\hat{\sigma}$ can be combined with threading to limit-support a as $(\gamma, \varepsilon) \to (1, 0)$.

Throughout the proof, let $\tilde{\varepsilon}_a = \sum_{\tilde{a} \neq a,b} \varepsilon_{a,\tilde{a}}$ be the probability that a player's action is recorded as something other than a or b when the player's action action is a, and let $\tilde{\varepsilon}_b = \sum_{\tilde{a} \neq a,b} \varepsilon_{b,\tilde{a}}$ be the probability that the action is recorded as something other than a or b when the actual action is b.

Claim OA.1 below shows that the steady-state share of good-standing players induced by $\hat{\sigma}$ converges to 1 uniformly over $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ as $\varepsilon \to 0$. For the remainder of the proof, we restrict attention to $\gamma \in [\underline{\gamma}, \overline{\gamma}]$. Claim OA.2 then shows that the incentives of good-standing players are satisfied when ε is sufficiently small. These two claims together complete the argument, as the incentives of bad-standing players are always satisfied since c is a strict best-response to b and (d, d) is a strict static equilibrium.

Claim OA.1. *For all $\delta > 0$, there is an $\overline{\varepsilon} > 0$ such that, whenever $\varepsilon_{x,x'} < \overline{\varepsilon}$ for all $x, x' \in A$, the steady states induced by $\hat{\sigma}$ satisfies $\mu^G > 1 - \delta$.*

Proof. Note that the inflow into good standing is $1-\gamma$, the share of newborn players. The outflow from good standing is the sum of $(1-\gamma)\mu^G$, the share of good-standing players who die in a given period, and $\gamma(\tilde{\varepsilon}_a\mu^G + \tilde{\varepsilon}_b(1-\mu^G))\mu^G$, the share of good-standing players who are recorded as playing an action other than a or b in a given period. In a steady state, these inflows and outflows must be equal, and setting the corresponding expressions equal to each other gives

$$\mu^G = \frac{1-\gamma}{1-\gamma+\gamma(\tilde{\varepsilon}_a\mu^G + \tilde{\varepsilon}_b(1-\mu^G))} \geq \frac{1-\gamma}{1-\gamma+\gamma\max\{\tilde{\varepsilon}_a,\tilde{\varepsilon}_b\}}.$$

The claim then follows since $\lim_{\varepsilon\to 0}\inf_{\gamma\in[\underline{\gamma},\overline{\gamma}]}(1-\gamma)/(1-\gamma+\gamma\max\{\tilde{\varepsilon}_a,\tilde{\varepsilon}_b\}) = 1$. ■

Claim OA.2. *For all $\delta > 0$, there is an $\overline{\varepsilon} > 0$ such that, whenever $\varepsilon_{x,x'} < \overline{\varepsilon}$ for all $x, x' \in A$, the incentives of good-standing players states are satisfied.*

Proof. We will use the facts that the value function of good-standing players, V^G, equals the average flow payoff in the population in a given period, so $\mu^G(\mu^G u(a,a) + (1-\mu^G)u(b,c)) + (1-\mu^G)(\mu^G u(c,b) + (1-\mu^G)u(d,d))$, and that the value function of bad-standing players is $V^B = \mu^G u(c,b) + (1-\mu^G)u(d,d)$.

When facing an opponent playing a, the expected payoff of a good-standing player from playing a is $(1-\gamma)u(a,a) + \gamma(1-\tilde{\varepsilon}_a)V^G + \tilde{\varepsilon}_a V^B$ while their expected payoff from playing b is $(1-\gamma)u(b,a) + \gamma(1-\tilde{\varepsilon}_b)V^G + \tilde{\varepsilon}_b V^B$. Thus, a good-standing player strictly prefers to play a rather than b precisely when

$$(1-\gamma)(u(a,a) - u(b,b)) > \gamma(\tilde{\varepsilon}_a - \tilde{\varepsilon}_b)(V^G - V^B). \tag{OA 2}$$

Moreover, the expected payoff of a good-standing player from playing action $x \notin \{a,b\}$ is $(1-\gamma)u(x,a) + \gamma(\varepsilon_{x,a} + \varepsilon_{x,b})V^G + \gamma(1-\varepsilon_{x,a}-\varepsilon_{x,b})V^B$. Thus, a good-standing player strictly prefers to play a rather than any $x \notin \{a,b\}$ precisely when

$$\frac{\gamma}{1-\gamma} > \max_{x\notin\{a,b\}} \frac{u(x,a) - u(a,a)}{(1-\tilde{\varepsilon}_a - \varepsilon_{x,a} - \varepsilon_{x,b})(\mu^G(u(a,a) - u(c,b)) + (1-\mu^G)(u(b,c) - u(d,d))}. \tag{OA 3}$$

Claim OA.1 implies that, as $\varepsilon \to 0$, the right-hand side of (OA 2) and the right-hand side of (OA 3) converge uniformly to 0 and $\max_{x \notin \{a,b\}}(u(x,a) - u(a,a))/(u(a,a) - u(c,b))$, respectively. From $u(a,a) > u(b,b)$ and (OA 1), we conclude that a good-standing player strictly prefers to match a with a for sufficiently small noise.

We now handle the incentives of a good-standing player to play b against an opponent who plays c. When facing an opponent playing c, the expected payoff of a good-standing player from playing a is $(1-\gamma)u(a,c) + \gamma(1-\tilde{\varepsilon}_a)V^G + \tilde{\varepsilon}_a V^B$ while their expected payoff from playing b is $(1-\gamma)u(b,c) + \gamma(1-\tilde{\varepsilon}_b)V^G + \tilde{\varepsilon}_b V^B$. Thus, a good-standing player strictly prefers to play b rather than a precisely when

$$(1-\gamma)(u(b,c) - u(a,c)) > \gamma(\tilde{\varepsilon}_b - \tilde{\varepsilon}_a)(V^G - V^B). \tag{OA 4}$$

Moreover, the expected payoff of a good-standing player from playing action $x \notin \{a,b\}$ is $(1-\gamma)u(x,c) + \gamma(\varepsilon_{x,a} + \varepsilon_{x,b})V^G + \gamma(1 - \varepsilon_{x,a} - \varepsilon_{x,b})V^B$. Thus, a good-standing player strictly prefers to play b rather than any $x \notin \{a,b\}$ precisely when

$$\frac{\gamma}{1-\gamma} > \max_{x \notin \{a,b\}} \frac{u(x,c) - u(b,c)}{(1 - \tilde{\varepsilon}_b - \varepsilon_{x,a} - \varepsilon_{x,b})(\mu^G(u(a,a) - u(c,b)) + (1-\mu^G)(u(b,c) - u(d,d))}. \tag{OA 5}$$

Claim OA.1 implies that as $\varepsilon \to 0$, the right-hand side of (OA 4) and the right-hand side of (OA 5) converge uniformly to 0 and $\max_{x \notin \{a,b\}}(u(x,c) - u(b,c))/(u(a,a) - u(c,b))$, respectively. From $u(b,c) > u(a,c)$ and (OA 1), we conclude that a good-standing player strictly prefers to play b rather than any other action against an opponent playing c for sufficiently small noise. ■

OA.3 Proofs of Lemmas for Theorem 5(ii)

OA.3.1 Proof of Lemma 12

Lemma 12. *There is a $D_J P_K S_1 D_\infty$ equilibrium with shares μ^{D_1}, μ^P, μ^S, and μ^{D_2} if and only if the following conditions hold:*

1. *Feasibility:*

$$\mu^{D_1} = 1 - \alpha(\gamma, 1-\varepsilon_D)^J,$$
$$\mu^P = \alpha(\gamma, 1-\varepsilon_D)^J(1-\beta(\gamma,\varepsilon,\mu^D)^K),$$
$$\mu^S = \alpha(\gamma, 1-\varepsilon_D)^J\beta(\gamma,\varepsilon,\mu^D)^K(1-\alpha(\gamma,\varepsilon_C)),$$
$$\mu^{D_2} = \alpha(\gamma, 1-\varepsilon_D)^J\beta(\gamma,\varepsilon,\mu^D)^K\alpha(\gamma,\varepsilon_C).$$

2. *Incentives:*

$$(C|C)_J: \frac{(1-\varepsilon_C-\varepsilon_D)\mu^D}{\varepsilon_C+(1-\varepsilon_C-\varepsilon_D)\mu^D}\left(\frac{\mu^S}{1-\mu^{D_1}}l + \frac{\mu^{D_2}}{\mu^{D_1}(1-\mu^{D_1})}(\mu^P-\mu^S g)\right) > g,$$
$$(D|D)_{J+K-1}: \frac{\gamma(1-\varepsilon_C-\varepsilon_D)((1-\alpha(\gamma,\varepsilon_C))\mu^D l + \alpha(\gamma,\varepsilon_C)(\mu^P-\mu^S g))}{1-\gamma+\gamma(\varepsilon_C+(1-\varepsilon_C-\varepsilon_D)\mu^D)} < l,$$
$$(C|D)_{J+K}\ (\text{if } \mu^S>0): \frac{\gamma(1-\varepsilon_C-\varepsilon_D)}{1-\gamma+\gamma\varepsilon_C}\left(\mu^P-\mu^S g-\mu^D l\right) > l.$$

We will derive the feasibility conditions and then derive the incentive conditions. The feasibility conditions of Lemma 12 are a consequence of the following lemma.

Lemma OA.1. *In a $D_J P_K S_1 D_\infty$ steady state with total share of defectors μ^D,*

$$\mu_k = \begin{cases} \alpha(\gamma,1-\varepsilon^D)^k(1-\alpha(\gamma,1-\varepsilon^D)) & \text{if } 0\le k\le J-1 \\ \alpha(\gamma,1-\varepsilon^D)^J\beta(\gamma,\varepsilon,\mu^D)^k(1-\beta(\gamma,\varepsilon,\mu^D)) & \text{if } J\le k\le J+K-1 \\ \alpha(\gamma,1-\varepsilon_D)^J\beta(\gamma,\varepsilon,\mu^D)^K(1-\alpha(\gamma,\varepsilon_C)) & \text{if } k=J+K \end{cases}.$$

To see why Lemma OA.1 implies the feasibility conditions of Lemma 12, note that

$$\mu^{D_1} = \sum_{k=0}^{J-1}\alpha(\gamma,1-\varepsilon_D)^k(1-\alpha(\gamma,1-\varepsilon_D)) = 1-\alpha(\gamma,1-\varepsilon_D)^J,$$
$$\mu^P = \sum_{k=J}^{J+K-1}\alpha(\gamma,1-\varepsilon_D)^J\beta(\gamma,\varepsilon,\mu^D)^k(1-\beta(\gamma,\varepsilon,\mu^D)) = \alpha(\gamma,1-\varepsilon_D)^J(1-\beta(\gamma,\varepsilon,\mu^D)^K),$$
$$\mu^S = \mu_{J+K} = \alpha(\gamma,1-\varepsilon_D)^J\beta(\gamma,\varepsilon,\mu^D)^K(1-\alpha(\gamma,\varepsilon_C)),$$

which also gives $\mu^{D_2} = 1-\mu^{D_1}-\mu^P-\mu^S = \alpha(\gamma,1-\varepsilon_D)^J\beta(\gamma,\varepsilon,\mu^D)^K\alpha(\gamma,\varepsilon_C)$.

Proof of Lemma OA.1. The inflow into score 0 is $1-\gamma$, while the outflow from score

0 is $(1-\gamma+\gamma(1-\varepsilon_D))\mu_0$. Setting these equal gives

$$\mu_0 = \frac{1-\gamma}{1-\gamma+\gamma(1-\varepsilon_D)} = 1-\alpha(\gamma, 1-\varepsilon_D).$$

Additionally, for every $0 < k < J$, both score k and score $k-1$ are defectors. Thus, the inflow into score k is $\gamma(1-\varepsilon_D)\mu_{k-1}$, while the outflow from score k is $(1-\gamma+\gamma(1-\varepsilon_D))\mu_k$. Setting these equal gives

$$\mu_k = \frac{\gamma(1-\varepsilon_D)}{1-\gamma+\gamma(1-\varepsilon_D)}\mu_{k-1} = \alpha(\gamma, 1-\varepsilon_D)\mu_{k-1}.$$

Combining these facts gives $\mu_k = \alpha(\gamma, 1-\varepsilon_D)^k(1-\alpha(\gamma, 1-\varepsilon_D))$ for $0 \leq k \leq J-1$.

Since record $J-1$ is a defector and record J is a preciprocator, the inflow into record J is $\gamma(1-\varepsilon_D)\mu_{J-1}$, while the outflow from record J is $(1-\gamma+\gamma(\varepsilon_C+(1-\varepsilon_C-\varepsilon_D)\mu^D))\mu_J$. Setting these equal and using the fact that $\mu_{J-1} = \alpha(\gamma, 1-\varepsilon_D)^{J-1}(1-\alpha(\gamma, 1-\varepsilon_D))$ gives

$$\begin{aligned}
\mu_J &= \alpha(\gamma, 1-\varepsilon_D)^{J-1}(1-\alpha(\gamma, 1-\varepsilon_D))\frac{\gamma(1-\varepsilon_D)}{1-\gamma+\gamma(\varepsilon_C+(1-\varepsilon_C-\varepsilon_D)\mu^D} \\
&= \alpha(\gamma, 1-\varepsilon_D)^J \frac{1-\gamma}{1-\gamma+\gamma(\varepsilon_C+(1-\varepsilon_C-\varepsilon_D)\mu^D} \\
&= \alpha(\gamma, 1-\varepsilon_D)^J(1-\beta(\gamma, \varepsilon, \mu^D)).
\end{aligned}$$

Additionally, for every $J < k < J+K$, both record k and record $k-1$ are preciprocators. Thus, the inflow into record k is $\gamma(\varepsilon_C+(1-\varepsilon_C-\varepsilon_D)\mu^D)\mu_{k-1}$, while the outflow from record k is $(1-\gamma+\gamma(\varepsilon_C+(1-\varepsilon_C-\varepsilon_D)\mu^D))\mu_k$. Setting these equal gives

$$\mu_k = \frac{\gamma(\varepsilon_C+(1-\varepsilon_C-\varepsilon_D)\mu^D)}{1-\gamma+\gamma(\varepsilon_C+(1-\varepsilon_C-\varepsilon_D)\mu^D)}\mu_{k-1} = \beta(\gamma, \varepsilon, \mu^D)\mu_{k-1}.$$

Combining these facts gives $\mu_k = \alpha(\gamma, 1-\varepsilon_D)^J\beta(\gamma, \varepsilon, \mu^D)^k(1-\beta(\gamma, \varepsilon, \mu^D))$ for $J \leq k \leq J+K-1$.

Since record $J+K-1$ is a preciprocator and record $J+K$ is a supercooperator, the inflow into record $J+K$ is $\gamma(\varepsilon_C + (1-\varepsilon_C-\varepsilon_D)\mu^D)\mu_{J+K-1}$, while the outflow is $(1-\gamma+\gamma\varepsilon_C))\mu_K$. Setting these equal and using the fact that $\mu_{J+K-1} = \alpha(\gamma, 1-\varepsilon_D)^J\beta(\gamma,\varepsilon,\mu^D)^{K-1}(1-\beta(\gamma,\varepsilon,\mu^D))$, we have

$$\begin{aligned}\mu_{J+K} &= \alpha(\gamma,1-\varepsilon_D)^J\frac{\gamma(\varepsilon_C+(1-\varepsilon_C-\varepsilon_D)\mu^D)}{1-\gamma+\gamma\varepsilon_C}\beta(\gamma,\varepsilon,\mu^D)^{K-1}(1-\beta(\gamma,\varepsilon,\mu^D))\\ &= \alpha(\gamma,1-\varepsilon_D)^J\beta(\gamma,\varepsilon,\mu^D)^K\frac{1-\gamma}{1-\gamma+\gamma\varepsilon_C}\\ &= \alpha(\gamma,1-\varepsilon_D)^J\beta(\gamma,\varepsilon,\mu^D)^K(1-\alpha(\gamma,\varepsilon_C)).\end{aligned}$$

■

Now we establish the incentive conditions in Lemma 12. We first handle the incentives of the score J preciprocator to play C against an opponent playing C. (When this incentive condition is satisfied, all other preciprocators play C against an opponent playing C.) Since V_J equals the average payoff in the population of players with score greater than J, we have

$$V_J = \frac{\mu^P}{1-\mu^{D_1}}\mu^C + \frac{\mu^S}{1-\mu^{D_1}}(\mu^C-\mu^D l) + \frac{\mu^{D_2}}{1-\mu^{D_1}}\mu^S(1+g).$$

Since the flow payoff to a preciprocator is μ^C, Lemma 7 along with the fact that $p_k^D = \varepsilon_C + (1-\varepsilon_C-\varepsilon_D)\mu^D$ for any preciprocator implies that a score J preciprocator plays C against C iff

$$\frac{1-\varepsilon_C-\varepsilon_D}{\varepsilon_C+(1-\varepsilon_C-\varepsilon_D)\mu^D}\left(\mu^C - \frac{\mu^P}{1-\mu^{D_1}}\mu^C - \frac{\mu^S}{1-\mu^{D_1}}(\mu^C-\mu^D l) - \frac{\mu^{D_2}}{1-\mu^{D_1}}\mu^S(1+g)\right) > g.$$

Since

$$\begin{aligned}&\mu^C - \frac{\mu^P}{1-\mu^{D_1}}\mu^C - \frac{\mu^S}{1-\mu^{D_1}}(\mu^C-\mu^D l) - \frac{\mu^{D_2}}{1-\mu^{D_1}}\mu^S(1+g)\\ =&\mu^D\left(\frac{\mu^S}{1-\mu^{D_1}}l + \frac{\mu^{D_2}}{\mu^{D_1}(1-\mu^{D_1})}(\mu^P-\mu^S g)\right),\end{aligned}$$

it follows that the $(C|C)_J$ constraint is equivalent to

$$\frac{(1-\varepsilon_C-\varepsilon_D)\mu^D}{\varepsilon_C+(1-\varepsilon_C-\varepsilon_D)\mu^D}\left(\frac{\mu^S}{1-\mu^{D_1}}l+\frac{\mu^{D_2}}{\mu^{D_1}(1-\mu^{D_1})}(\mu^P-\mu^S g)\right)>g.$$

To handle the incentives of a score $J+K$ supercooperator, note that

$$V_{J+K}=(1-\gamma)(\mu^C-\mu^D l)+\gamma(1-\varepsilon_C)V_K+\gamma\varepsilon_C V_{J+K+1}.$$

Combining this with the fact that $V_k=\mu^S(1+g)$ for all $k>K+J$ gives

$$V_{J+K}=(1-\alpha(\gamma,\varepsilon_C))(\mu^C-\mu^D l)+\alpha(\gamma,\varepsilon_C)\mu^S(1+g). \tag{OA 6}$$

Thus, we have

$$\frac{\gamma(1-\varepsilon_C-\varepsilon_D)}{1-\gamma}(V_{J+K}-V_{J+K+1})=\frac{\gamma(1-\varepsilon_C-\varepsilon_D)}{1-\gamma+\gamma\varepsilon_C}(\mu^P-\mu^S g-\mu^D l),$$

from which the $(C|D)_{J+K}$ constraint in Lemma 12 immediately follows.

Finally, we show that a record $J+K-1$ preciprocator prefers to play D against an opponent playing D. (This implies that all other preciprocators play D against an opponent playing D.) Note that

$$V_{J+K-1}=(1-\gamma)\mu^C+\gamma(1-\varepsilon_C-(1-\varepsilon_C-\varepsilon_D)\mu^D)V_{K-1}+\gamma(\varepsilon_C+(1-\varepsilon_C-\varepsilon_D)\mu^D)V_{J+K},$$

so

$$\frac{\gamma(1-\varepsilon_C-\varepsilon_D)}{1-\gamma}(V_{J+K-1}-V_{J+K})=\frac{\gamma(1-\varepsilon_C-\varepsilon_D)}{1-\gamma+\gamma(\varepsilon_C+(1-\varepsilon_C-\varepsilon_D)\mu^D)}(\mu^C-V_{J+K}).$$

Combining this with the expression for V_{J+K} in Equation OA 6 gives

$$\frac{\gamma(1-\varepsilon_C-\varepsilon_D)}{1-\gamma}(V_{J+K-1}-V_{J+K})=\frac{\gamma(1-\varepsilon_C-\varepsilon_D)((1-\alpha(\gamma,\varepsilon_C))\mu^D l+\alpha(\gamma,\varepsilon_C)(\mu^P-\mu^S g))}{1-\gamma+\gamma(\varepsilon_C+(1-\varepsilon_C-\varepsilon_D)\mu^D)},$$

which implies the form of the $(D|D)_{J+K-1}$ constraint in Lemma 12.

OA.3.2 Proof of Lemma 13

Lemma 13. *There are $0 < \underline{\gamma} < \overline{\gamma} < 1$ and $\overline{\varepsilon} > 0$ such that, for all $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ and $\varepsilon_C, \varepsilon_D < \overline{\varepsilon}$, there is a $D_J P_K S_1 D_\infty$ strategy with a steady state whose shares satisfy $|\mu_1^D - \delta|, |\mu^P - \overline{\mu}^P|, |\mu^S - \overline{\mu}^S| \leq \eta$, and are such that the $(C|D)_{J+K}$ constraint in Lemma 12 is satisfied.*

Let $J(\gamma, \delta) = \lceil \ln(1-\delta)/\ln(\gamma) \rceil$ be the smallest integer greater than $\ln(1-\delta)/\ln(\gamma)$. Let $K(\gamma, \delta) = \left\lceil (\ln(\gamma^{J(\gamma,\delta)} - \overline{\mu}^P) - \ln(\gamma^{J(\gamma,\delta)}))/\ln(\beta(\gamma, 0, \delta)) \right\rceil$. Let $\overline{\gamma} \in ((1+\delta)/2, 1)$ be such that

$$
\begin{aligned}
|\overline{\gamma}^{J(\overline{\gamma},\delta)} - (1-\delta)| &\leq \frac{\eta}{6}, \\
\left|\overline{\gamma}^{J(\overline{\gamma},\delta)}(1 - \beta(\overline{\gamma}, 0, \delta)^{K(\overline{\gamma},\delta)}) - \overline{\mu}^P\right| &\leq \frac{\eta}{6}, \\
\left|\overline{\gamma}^{J(\overline{\gamma},\delta)} \left(1 - \beta\left(\overline{\gamma}, 0, \delta + 2(1-\overline{\gamma})\right)^{K(\overline{\gamma},\delta)}\right) - \overline{\mu}^P\right| &\leq \frac{\eta}{6}, \\
\frac{\overline{\gamma}}{1-\overline{\gamma}}(\overline{\mu}^P - \eta - (\overline{\mu}^S + \eta)g - (\delta + 2\eta)l) &> l.
\end{aligned}
\tag{OA 7}
$$

To see that such a $\overline{\gamma}$ exists, note that $\lim_{\gamma \to 1} \gamma^{J(\gamma,\delta)} = 1-\delta$ and $\lim_{\gamma \to 1} \beta(\gamma, 0, \delta)^{K(\gamma,\delta)} = 1 - \overline{\mu}^P/(1-\delta)$, so $\lim_{\gamma \to 1} \gamma^{J(\gamma,\delta)}(1 - \beta(\gamma, 0, \delta)^{K(\gamma,\delta)}) = \overline{\mu}^P$. Additionally, since $\overline{\mu}^P - \eta - (\overline{\mu}^S + \eta)g - (\delta + 2\eta)l > 0$, the left-hand side of the fourth inequality approaches infinity as $\gamma \to 1$. The argument for the third inequality is a little more involved. Let $K'(\gamma, \delta) = \left\lceil (\ln(\gamma^{J(\gamma,\delta)} - \overline{\mu}^P) - \ln(\gamma^{J(\gamma,\delta)}))/\ln(\beta(\gamma, 0, \delta + 2(1-\gamma)) \right\rceil$. It can be shown that $\lim_{\gamma \to 1} K(\gamma, \delta)/K'(\gamma, \delta) = 1$. Moreover, $\lim_{\gamma \to 1} \beta(\gamma, 0, \delta + 2(1-\gamma))^{K'(\gamma,\delta)} = 1 - \overline{\mu}^P/(1-\delta)$, so it follows that $\lim_{\gamma \to 1} \beta(\gamma, 0, \delta + 2(1-\gamma))^{K(\gamma,\delta)} = \lim_{\gamma \to 1} (\beta(\gamma, 0, \delta + 2(1-\gamma))^{K'(\gamma,\delta)})^{K(\gamma,\delta)/K'(\gamma,\delta)} = 1 - \overline{\mu}^P/(1-\delta)$. Combining this with $\lim_{\gamma \to 1} \gamma^{J(\gamma,\delta)} = 1 - \delta$ gives $\lim_{\gamma \to 1} \gamma^{J(\gamma,\delta)} \left(1 - \beta\left(\gamma, 0, \delta + 2(1-\gamma)\right)^{K(\gamma,\delta)}\right) = \overline{\mu}^P$.

Let $\overline{J} = J(\overline{\gamma}, \delta)$ and $\overline{K} = K(\overline{\gamma}, \delta)$. There exists some $\underline{\gamma} \in ((1+\delta)/2, \overline{\gamma})$ such that $\overline{J} - 1 \leq \ln(1-\delta)/\ln(\gamma) \leq \overline{J}$ for all $\gamma \in [\underline{\gamma}, \overline{\gamma}]$. Moreover, continuity, combined with the inequalities in (OA 7), implies that this $\underline{\gamma}$ can be chosen along with some $\overline{\varepsilon} > 0$ such that

$$\begin{aligned}
|\alpha(\gamma, 1-\varepsilon_D)^{\overline{J}} - (1-\delta)| &\leq \frac{\eta}{3}, \\
\left|\alpha(\gamma, 1-\varepsilon_D)^{\overline{J}}(1-\beta(\gamma,\varepsilon,\delta)^{\overline{K}}) - \overline{\mu}^P\right| &\leq \frac{\eta}{3}, \\
\left|\alpha(\gamma, 1-\varepsilon_D)^{\overline{J}}\left(1-\beta\left(\gamma,\varepsilon,\delta+2(1-\gamma)\right)^{\overline{K}}\right) - \overline{\mu}^P\right| &\leq \frac{\eta}{3}, \\
\frac{\gamma(1-\varepsilon_C-\varepsilon_D)}{1-\gamma+\gamma\varepsilon_C}(\overline{\mu}^P - \eta - (\overline{\mu}^S+\eta)g - (\delta+2\eta)l) &> l,
\end{aligned} \tag{OA 8}$$

for all $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ and $\varepsilon_C, \varepsilon_D < \overline{\varepsilon}$.

Since $\mu^{D_2} \leq \alpha(\gamma, \varepsilon_C)$ and $\alpha(\gamma, \varepsilon_C) \to 0$ as $\varepsilon_C \to 0$ uniformly over $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, we can take $\overline{\varepsilon}$ to be such that $\mu^{D_2} \leq \min\{\eta/3, (1-\gamma)/2\}$ for all $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ and $\varepsilon_C, \varepsilon_D < \overline{\varepsilon}$. Moreover, as $\overline{J} - 1 \leq \ln(1-\delta)/\ln(\gamma) \leq \overline{J}$, it follows that $\gamma^{\overline{J}} \in [\gamma(1-\delta), 1-\delta]$ for all $\gamma \in [\underline{\gamma}, \overline{\gamma}]$. Because $\alpha(\gamma, 1-\varepsilon_D) \leq \gamma$ and $\alpha(\gamma, 1-\varepsilon_D) \to \gamma$ as $\varepsilon_D \to 0$ uniformly over $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, we can take $\overline{\varepsilon}$ to be such that $\mu^{D_1} = 1 - \alpha(\gamma, 1-\varepsilon_D)^{\overline{J}} \in [\delta, \delta + 3(1-\gamma)/2]$ for all $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ and $\varepsilon_C, \varepsilon_D < \overline{\varepsilon}$. Thus, $\mu^D \in [\delta, \delta + 2(1-\gamma)]$ for all $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ and $\varepsilon_C, \varepsilon_D < \overline{\varepsilon}$. As $\beta(\gamma, \varepsilon, \mu^D)$ is increasing in μ^D, the first three inequalities in (OA 8) imply that, for all $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ and $\varepsilon_C, \varepsilon_D < \overline{\varepsilon}$, there are feasible steady states with $|\mu^{D_1} - \delta|, |\mu^P - \overline{\mu}^P|, \mu^{D_2} \leq \eta/3$. Additionally, since $\overline{\mu}^S = 1 - \delta - \overline{\mu}^P$ and $\mu^S = 1 - \mu^{D_1} - \mu^P - \mu^{D_2}$, it follows that all such steady states must have $|\mu^S - \overline{\mu}^S| \leq \eta$. Finally, note that these facts, along with the fourth inequality in (OA 8), imply that the $(C|D)_{J+K}$ constraint in Lemma 12 is satisfied.