Appendix

A.1 Proof of Theorem 1

We first prove Theorem 1(i).

Without loss, relabel records so that two players with different ages can never share the same record. Let \( R(t) \) be the set of feasible records for an age-\( t \) player, and fix a pairwise-public strategy \( \sigma \). The proof relies on the following lemma.

**Lemma 4.** If records are finite-partitional, there exists a family of finite subsets of \( R \), \( \{L(t, \eta)\}_{t \in \mathbb{N}, \eta > 0} \), such that

1. \( L(t, \eta) \subset R(t) \) for all \( t \in \mathbb{N}, \eta > 0 \),
2. For any \( \mu \in \Delta(R) \), \( \sum_{r \in L(0, \eta)} f_\sigma(\mu)[r] \geq (1 - \eta)(1 - \gamma) \) for all \( \eta > 0 \), and
3. For any \( \mu \in \Delta(R) \) and \( t > 0 \), if \( \sum_{r \in L(t-1, \eta)} \mu_r \geq (1 - \eta)(1 - \gamma)\gamma^{t-1} \) for all \( \eta > 0 \), then \( \sum_{r \in L(t, \eta)} f_\sigma(\mu)[r] \geq (1 - \eta)(1 - \gamma)\gamma^t \) for all \( \eta > 0 \).

**Proof.** We construct the \( \{L(t, \eta)\} \) by iteratively defining subfamilies of subsets of \( R \) that satisfy the necessary properties. First, take \( L(0, \eta) = \{0\} \) for all \( \eta > 0 \). Conditions 1 and 2 are satisfied since \( R(0) = \{0\} \) and \( f_\sigma(\mu)[0] = 1 - \gamma \) for every \( \mu \in \Delta(R) \).

Fix some \( t \) and take the subfamily of subsets corresponding to \( t - 1 \), that is \( \{L(t - 1, \eta)\}_{\eta > 0} \). For every \( \eta > 0 \), consider the set of records \( L(t - 1, \eta/2) \). Let \( \lambda \in (0, 1) \) be such that \( \lambda \geq (1 - \eta)/(1 - \eta/2) \). For any record \( r \in L(t - 1, \eta/2) \), opposing record class \( R_m \), and action profile \( (a, a') \in A^2 \), we can identify a finite set of “successor records” \( S(r, m, a, a') \) such that a record \( r \) player who plays \( a \) against an opponent in class \( R_m \) playing \( a' \) moves to a record in \( S(r, m, a, a') \) with probability greater than \( \lambda \), i.e. \( \sum_{r'' \in S(r, m, a, a')} \rho(r, r', a, a')[r''] \geq \lambda \) for all \( r' \in R_m \). Let \( L(t, \eta) = \bigcup_{r \in L(t-1, \eta/2)} \bigcup_{m \in \{1, \ldots, M(r)\}} \bigcup_{(a, a') \in A^2} S(r, m, a, a') \). Note that \( L(t, \eta) \) is finite and does not depend on \( \mu \). By construction, the probability that a surviving player with record in \( L(t - 1, \eta/2) \) has a next-period record in \( L(t, \eta) \) exceeds \( \lambda \). For
any $\mu \in \Delta(R)$, it then follows that $\sum_{r \in L(t-1,\eta/2)} \mu_r \geq (1 - \eta/2)(1 - \gamma)\gamma^{t-1}$ implies $\sum_{r \in L(t,\eta)} f_\sigma(\mu)[r] \geq \lambda(1 - \eta/2)(1 - \gamma)\gamma^{t-1} \geq (1 - \eta)(1 - \gamma)\gamma^t$.

The next corollary is an immediate consequence of Properties 2 and 3 of Lemma 4.

**Corollary 4.** For every $\mu \in \Delta(R)$ and $\eta > 0$, we have $\sum_{r \in L(t,\eta)} f_\sigma^{t}(\mu)[r] \geq (1 - \eta)(1 - \gamma)\gamma^t$ for all $t' > t$, where $f_\sigma^{t'}$ denotes the $t'$th iterate of the update map $f_\sigma$.

Fix a family $\{L(t,\eta)\}_{t \in \mathbb{N}, \eta > 0}$, satisfying the three properties in Lemma 4 and define $\bar{M}$, a subset of $\Delta(R)$, by

$$
\bar{M} = \left\{ \mu \in \Delta(R) : \sum_{r \in R(t)} \mu_r = (1 - \gamma)\gamma^t \text{ and } \sum_{r \in L(t,\eta)} \mu_r \geq (1 - \eta)(1 - \gamma)\gamma^t \forall t \in \mathbb{N}, \eta > 0 \right\}.
$$

Note that $\bar{M}$ is convex and, by Corollary 4, must contain every steady-state distribution $\mu$. The next lemma uses Corollary 4 to show that $\bar{M}$ is non-empty.

**Lemma 5.** There exists $\mu \in \Delta(R)$ satisfying $\sum_{r \in R(t)} \mu_r = (1 - \gamma)\gamma^t$ and $\sum_{r \in L(t,\eta)} \mu_r \geq (1 - \eta)(1 - \gamma)\gamma^t$ for every $t \in \mathbb{N}, \eta > 0$.

**Proof.** Consider an arbitrary $\mu \in \Delta(R)$. Set $\mu^0 = \mu$, and, for every non-zero $i \in \mathbb{N}$, set $\mu^i = f_\sigma(\mu^{i-1})$. Since $R$ is countable, a standard diagonalization argument implies that there exists some $\bar{\mu} \in [0, 1]^R$ and some subsequence $\{\mu^{i_j}\}_{j \in \mathbb{N}}$ such that $\lim_{j \to \infty} \mu^{i_j} = \bar{\mu}$, for all $r \in R$.

For a given $t \in \mathbb{N}$, Corollary 4 implies that $\sum_{r \in L(t,\eta)} \mu^{i_j}_r \geq (1 - \eta)(1 - \gamma)\gamma^t$ for all $\eta > 0$ and all sufficiently high $j \in \mathbb{N}$, so

$$
\sum_{r \in L(t,\eta)} \bar{\mu}_r \geq (1 - \eta)(1 - \gamma)\gamma^t.
$$

Moreover, for each $t \in \mathbb{N}$, $\sum_{r \in R(t)} \mu^{i_j}_r = (1 - \gamma)\gamma^t$ for all $j \in \mathbb{N}$, so $\sum_{r \in R(t)} \bar{\mu}_r \leq (1 - \gamma)\gamma^t$. Since (1) holds for all $\eta \in (0, 1)$, this implies that $\sum_{r \in R(t)} \bar{\mu}_r = (1 - \gamma)\gamma^t$, which together with (1) implies that $\bar{\mu} \in \bar{M}$.

\(\blacksquare\)
The following three claims imply that \( f_\sigma \) has a fixed point in \( \bar{M} \),\(^{22}\) which completes the proof of parts 1 and 2 of Theorem 1.

**Claim 1.** \( \bar{M} \) is compact in the sup norm topology.

**Claim 2.** \( f_\sigma \) maps \( \bar{M} \) to itself.

**Claim 3.** \( f_\sigma \) is continuous in the sup norm topology.

*Proof of Claim 1.* Since \( \bar{M} \) is a metric space under the sup norm topology, it suffices to show that \( \bar{M} \) is sequentially compact. Consider a sequence \( \{\mu^i\}_{i \in \mathbb{N}} \) of \( \mu^i \in \bar{M} \). A similar argument to the proof of Lemma 5 shows that there exists some \( \tilde{\mu} \in \bar{M} \) and some subsequence \( \{\mu^j\}_{j \in \mathbb{N}} \) such that \( \lim_{j \to \infty} \mu^j = \tilde{\mu} \) for all \( r \in R \).

Here we show that \( \lim_{j \to \infty} \mu^j = \tilde{\mu} \). For a given \( \eta > 0 \), there is a finite subset of records \( L(\eta/2) \subset R \) such that \( \sum_{r \in L(\eta)} \mu_r > 1 - \eta/2 \) for every \( \mu \in \bar{M} \). Thus, \( |\mu^j - \tilde{\mu}| < \eta/2 \) for all \( r \notin L(\eta/2) \) for all \( j \in \mathbb{N} \). Now let \( J \in \mathbb{N} \) be such that \( |\mu^j - \tilde{\mu}| < \eta/2 \) for all \( r \in L(\eta/2) \) whenever \( j > J \). Then \( \sup_{r \in R} |\mu^j - \tilde{\mu}| < \eta \) for all \( j > J \). \( \blacksquare \)

*Proof of Claim 2.* For any \( \mu \in \bar{M} \), Properties 2 and 3 of Lemma 4 imply that \( \sum_{r \in L(t,\eta)} f_\sigma(\mu)[r] \geq (1 - \eta)(1 - \gamma)\gamma^t \) for all \( t \in \mathbb{N}, \eta > 0 \). Furthermore, \( f_\sigma(\mu)[0] = 1 - \gamma \), and for all \( t > 0 \), \( \gamma \sum_{r \in R(t-1)} \mu_r = \sum_{r \in R(t)} f_\sigma(\mu)[r] \), so \( \sum_{r \in R(t-1)} \mu_r = (1 - \gamma)\gamma^{t-1} \) gives \( \sum_{r \in R(t)} f_\sigma(\mu)[r] = (1 - \gamma)\gamma^t \). \( \blacksquare \)

*Proof of Claim 3.* Consider a sequence \( \{\mu^i\}_{i \in \mathbb{N}} \) of \( \mu^i \in \bar{M} \) with \( \lim_{i \to \infty} \mu^i = \tilde{\mu} \in \bar{M} \). We will show that \( \lim_{i \to \infty} f_\sigma(\mu^i) = f_\sigma(\tilde{\mu}) \).

For any \( \eta > 0 \), there is a finite subset of records \( L(\eta/4) \subset R \) such that \( \sum_{r \in L(\eta/4)} \mu_r > 1 - \eta/4 \) for every \( \mu \in \bar{M} \). By Claim 2, \( f_\sigma(\mu) \in \bar{M} \) for every \( \mu \in \bar{M} \). The combination of these facts means that it suffices to show that \( \lim_{i \to \infty} f_\sigma(\mu^i)[r] = f_\sigma(\tilde{\mu})[r] \) for all \( r \in R \) to establish \( \lim_{i \to \infty} f_\sigma(\mu^i) = f_\sigma(\tilde{\mu}) \). Additionally, since \( f_\sigma(\mu)[0] = 1 - \gamma \) is constant across \( \mu \in \Delta(R) \), we need only consider the case where \( r \neq 0 \).

\(^{22}\)This follows from Corollary 17.56 (page 583) of Aliprantis and Border (2006), and noting that every normed space is a locally convex Hausdorff space.
For this case,

\[ f_\sigma(\mu^i)[r] = \gamma \sum_{(r', r'') \in R^2} \mu_{r', r''} \phi(r', r'')[r], \]

and

\[ f_\sigma(\bar{\mu})[r] = \gamma \sum_{(r', r'') \in R^2} \bar{\mu}_{r', r''} \phi(r', r'')[r]. \]

Because \( \sum_{r \in L(\eta/4)} \mu_r > 1 - \eta/4 \) for every \( \mu \in \bar{M}, \gamma \in (0, 1) \), and \( 0 \leq \phi(r', r'')[r] \leq 1 \) for all \( r', r'' \in R \), it follows that

\[
|f_\sigma(\mu^i)[r] - f_\sigma(\bar{\mu})[r]| \leq \gamma \sum_{(r', r'') \in L(\eta/4)^2} |(\mu_{r', r''} - \bar{\mu}_{r', r''})\phi(r', r'')[r]|
\]

\[
+ \gamma \sum_{(r', r'') \notin L(\eta/4)^2} |(\mu_{r', r''} - \bar{\mu}_{r', r''})\phi(r', r'')[r]|
\]

\[
< \sum_{(r', r'') \in L(\eta/4)^2} |\mu_{r', r''} - \bar{\mu}_{r', r''}| + \frac{1}{2} \eta.
\]

Since \( \lim_{i \to \infty} \mu^i = \bar{\mu} \), there exists some \( I \in \mathbb{N} \) such that \( \sum_{(r', r'') \in L(\eta/4)^2} |\mu_{r', r''} - \bar{\mu}_{r', r''}| < \eta/2 \) for all \( i > I \), which gives \( |f_\sigma(\mu^i)[r] - f_\sigma(\bar{\mu})[r]| < \eta \) for all \( i > I \). We thus conclude that \( \lim_{i \to \infty} f_\sigma(\mu^i)[r] = f_\sigma(\bar{\mu})[r] \). \hfill \blacksquare

We now prove Theorem 1(ii) by showing that no steady state exists when \( \gamma > 1/2 \) for the interdependent record system with \( R = \mathbb{N} \) and \( \rho(r, r') = \max\{r, r\} + 1 \). To see this, suppose toward a contradiction that \( \mu \) is a steady state. Let \( r^* \) be the smallest record \( r \) such that \( \sum_{r'=r}^{\infty} \mu_{r'} < 2 - 1/\gamma \), and let \( \mu_* = \sum_{r=r^*}^{\infty} \mu_r < 2 - 1/\gamma \). Note that \( \mu_* > 0 \), as a player’s record is no less than their age, so for any record threshold there is a positive measure of players whose records exceed the threshold.

Note that every surviving player with record \( r \geq r^* \) retains a record higher than \( r^* \), and at least fraction \( \mu_* \) of the surviving players with record \( r < r^* \) obtain a record higher than \( r^* \) (since this is the fraction of players with record \( r < r^* \) that match with
a player with record \( r \geq r^* \). Hence,

\[
\sum_{r=r^*}^{\infty} f(r) \geq \gamma \mu_* + \gamma (1 - \mu_*) \mu_* > \mu_*,
\]

where the second inequality comes from \( 0 < \mu_* < 2 - 1/\gamma \). But in a steady state, \( \sum_{r=r^*}^{\infty} f(r) = \mu_* \), a contradiction.

## A.2 Proof of Theorem 2

### A.2.1 Proof of Theorem 2(i)

Let \( M \) be a positive integer such that \((u(a, a) - u(b, b))M > \max_{a'}\{\max\{u(a', a) - u(a, a), u(a', b) - u(b, b)\}\}\}. We show that, with this choice of \( M \), action \( a \) can be limit-supported by the cyclic strategies defined in Section 3, which we denote by \( \sigma^* \).23

Let \( \tilde{\varepsilon}_{(a,a)} = \sum_{(\tilde{a},\tilde{a}') \neq (a,a), (b,b)} \varepsilon_{(a,a), (\tilde{a},\tilde{a}')} \) be the probability that the stage-game outcome is recorded as something other than \((a,a)\) or \((b,b)\) when the actual outcome is \((a,a)\), \( \tilde{\varepsilon}_{(b,b)} = \sum_{(\tilde{a},\tilde{a}') \neq (a,a), (b,b)} \varepsilon_{(b,b), (\tilde{a},\tilde{a}')} \) be the probability that the outcome is recorded as something other than \((a,a)\) or \((b,b)\) when the actual outcome is \((b,b)\), and \( \tilde{\varepsilon}_{(b,b)} = \sum_{(\tilde{a},\tilde{a}') \neq (b,b), (\tilde{a},\tilde{a}')} \varepsilon_{(b,b), (\tilde{a},\tilde{a}')} \) be the probability that the outcome is recorded as something other than \((b,b)\) when the actual outcome is \((b,b)\).

Consider a steady state \( \mu(\gamma, \varepsilon) \) for parameters \((\gamma, \varepsilon)\). Let \( \mu^G(\gamma, \varepsilon) \) be the corresponding share of good-standing players. Similarly, for \( i \in \{0, ..., M - 1\} \), let \( \mu^B_i(\gamma, \varepsilon) \) be the share of bad-standing players who have accumulated \( i \) \((b,b)\) profiles since last entering bad standing. We show that the unique limit point of any sequence of steady-state shares \((\mu^G(\gamma, \varepsilon), \mu^B_0(\gamma, \varepsilon), ..., \mu^B_{M-1}(\gamma, \varepsilon))\) as \((\gamma, \varepsilon) \to (1, 0)\) is \((\bar{\mu}^G, \bar{\mu}^B_0, ..., \bar{\mu}^B_{M-1}) = (1, 0, ..., 0)\). This implies that \( \lim_{(\gamma, \varepsilon) \to (1, 0)} \mu^G(\gamma, \varepsilon) = 1 \), so the share of good-standing players converges to 1 in the \((\gamma, \varepsilon) \to (1, 0)\) limit. Consequently, the population share of action \( a \) also converges to 1.

---

23Note that the strategy \( \sigma^* \) does not depend on \((\gamma, \varepsilon)\).
Let \((\tilde{\mu}^G, \tilde{\mu}^B_0, \ldots, \tilde{\mu}^B_{M-1})\) be a limit point of a sequence of steady-state shares as \((\gamma, \varepsilon) \to (1, 0)\). The inflow into \(B_0\), the first phase of bad-standing, is \(\gamma(1 - \tilde{\varepsilon}_{(b,b)} - (1 - \tilde{\varepsilon}_{(a,a)} - \tilde{\varepsilon}_{(b,b)})\mu^G(\gamma, \varepsilon))\mu^G(\gamma, \varepsilon)\), which is the share of good-standing players that move into bad-standing in a given period. The outflow from \(B_0\) is the sum of \((1 - \gamma)\mu^B_0(\gamma, \varepsilon)\), the share of players in phase \(B_0\) who die in a given period, and \(\gamma(1 - \tilde{\varepsilon}_{(b,b)})\mu^B_0(\gamma, \varepsilon)\), the share of players in phase \(B_0\) who move into phase \(B_1\) in a given period. Thus, in a steady state, \(\gamma(1 - \tilde{\varepsilon}_{(b,b)} - (1 - \tilde{\varepsilon}_{(a,a)} - \tilde{\varepsilon}_{(b,b)})\mu^G(\gamma, \varepsilon))\mu^G(\gamma, \varepsilon) = (1 - \gamma\tilde{\varepsilon}_{(b,b)})\mu^B_0(\gamma, \varepsilon)\). Taking the limit of this equation as \((\gamma, \varepsilon) \to (1, 0)\) gives \(\tilde{\mu}^B_0 = 0\). Likewise, equating the inflow and outflows of phase \(B_i\) for \(0 < i < M\) gives \(\gamma(1 - \tilde{\varepsilon}_{(b,b)})\mu^{B_{i-1}}(\gamma, \varepsilon) = (1 - \gamma\tilde{\varepsilon}_{(b,b)})\mu^{B_i}(\gamma, \varepsilon)\), and taking the limit of this equation as \((\gamma, \varepsilon) \to (1, 0)\) shows that \(\tilde{\mu}^{B_i} = \tilde{\mu}^{B_{i-1}}\). Combining this with \(\tilde{\mu}^B_0 = 0\) gives \(\tilde{\mu}^{B_i} = 0\) for all \(i \in \{0, \ldots, M - 1\}\). Since the good-standing population share and bad-standing population shares always sum to 1, it follows that \(\tilde{\mu}^G = 1\).

We now show that \((\sigma^*, \mu(\gamma, \varepsilon))\) is a strict equilibrium when \(\gamma\) is sufficiently close to 1 and \(\varepsilon\) is sufficiently close to 0. For \(0 \leq i < M - 1\), the value functions in the bad-standing phase \(B_i\) and the subsequent bad-standing phase \(B_{i+1}\) satisfy

\[
V^{B_i} = (1 - \gamma)u(b, b) + \gamma\tilde{\varepsilon}_{(b,b)}V^{B_i} + \gamma(1 - \tilde{\varepsilon}_{(b,b)})V^{B_{i+1}}. \tag{2}
\]

Similarly the value functions in the final bad-standing phase \(B_{M-1}\) and the good-standing phase \(G\) are linked by

\[
V^{B_{M-1}} = (1 - \gamma)u(b, b) + \gamma\tilde{\varepsilon}_{(b,b)}V^{B_{M-1}} + \gamma(1 - \tilde{\varepsilon}_{(b,b)})V^G. \tag{3}
\]

Combining \(\lim_{(\gamma, \varepsilon) \to (1, 0)} \mu^G(\gamma, \varepsilon) = 1\) with \(V^G = \mu^G(\gamma, \varepsilon)^2u(a, a) + (1 - \mu^G(\gamma, \varepsilon)^2)u(b, b)\) shows that \(\lim_{(\gamma, \varepsilon) \to (1, 0)} V^G = u(a, a)\). Taking the limits of these equations as \((\gamma, \varepsilon) \to (1, 0)\) gives \(\lim_{(\gamma, \varepsilon) \to (1, 0)} V^{B_i} = \lim_{(\gamma, \varepsilon) \to (1, 0)} V^G = u(a, a)\) for all \(i \in \{0, \ldots, M - 1\}\).

A player in bad-standing phase \(i\) where \(0 \leq i < M - 1\) strictly prefers to play \(b\) against \(b\) when \((1 - \gamma)u(b, b) + \gamma\tilde{\varepsilon}_{(b,b)}V^{B_i} + \gamma(1 - \tilde{\varepsilon}_{(b,b)})V^{B_{i+1}} > (1 - \gamma)u(a', b) + \gamma(1 -
\[ \varepsilon_{(a',b),(b,b)} V^{B_i} + \gamma \varepsilon_{(a',b),(b,b)} V^{B_{i+1}} \text{ holds for } a' \neq b. \] Manipulating this gives \((1 - \tilde{\varepsilon}_{(b,b)} - \varepsilon_{(a',b),(b,b)} \gamma (V^{B_{i+1}} - V^{B_i})/(1 - \gamma) > u(a', b) - u(b, b).\) Equation 2 can be rewritten as

\[
\frac{\gamma}{1 - \gamma} (V^{B_{i+1}} - V^{B_i}) = \frac{\gamma}{1 - \gamma \tilde{\varepsilon}_{(b,b)}} (V^{B_{i+1}} - u(b, b)),
\]

so we obtain \(\lim_{(\gamma, \varepsilon) \to (1,0)} (1 - \tilde{\varepsilon}_{(b,b)} - \varepsilon_{(a',b),(b,b)} \gamma (V^{B_{i+1}} - V^{B_i})/(1 - \gamma) = u(a, a) - u(b, b).\) Since \(\max_{a'} u(a', b) < u(a, a),\) it follows that the incentives of players in bad-standing phase \(i\) are satisfied for \((\gamma, \varepsilon)\) sufficiently close to \((1,0).\)

An almost identical argument shows that the incentives of players in bad-standing phase \(M - 1\) are satisfied for \((\gamma, \varepsilon)\) sufficiently close to \((1,0).\) Thus, all that remains is to show that the incentives of players in good-standing are satisfied in the limit. A good-standing player has strict incentives to play \(a\) against \(a\) when \((1 - \gamma) u(a, a) + \gamma (1 - \tilde{\varepsilon}_{(a,a)}) V^G + \gamma \tilde{\varepsilon}_{(a,a)} V^{B_0} > (1 - \gamma) u(a', a) + \gamma (\varepsilon_{(a',a),(a,a)} + \varepsilon_{(a',a),(b,b)}) V^G + \gamma (1 - \varepsilon_{(a',a),(a,a)} - \varepsilon_{(a',a),(b,b)}) V^{B_0}\) holds for \(a' \neq a.\) Manipulating this gives \((1 - \tilde{\varepsilon}_{(a,a)} - \varepsilon_{(a',a),(a,a)} - \varepsilon_{(a',a),(b,b)}) \gamma (V^G - V^{B_0})/(1 - \gamma) > u(a', a) - u(a, a).\) Similarly, a good-standing player has strict incentives to play \(b\) against \(b\) when \((1 - \gamma) u(b, b) + \gamma (1 - \tilde{\varepsilon}_{(b,b)}) V^G + \gamma \tilde{\varepsilon}_{(b,b)} V^{B_0} > (1 - \gamma) u(a', b) + \gamma (\varepsilon_{(a',b),(a,a)} + \varepsilon_{(a',b),(b,b)}) V^G + \gamma (1 - \varepsilon_{(a',b),(a,a)} - \varepsilon_{(a',b),(b,b)}) V^{B_0}\) holds for \(a' \neq b.\) Manipulating this gives \((1 - \tilde{\varepsilon}_{(b,b)} - \varepsilon_{(a',b),(a,a)} - \varepsilon_{(a',b),(b,b)}) \gamma (V^G - V^{B_0})/(1 - \gamma) > u(a', b) - u(b, b).\) Combining Equations 2 and 3 gives

\[ \frac{\gamma}{1 - \gamma} (V^G - V^{B_0}) = \frac{1 - \left(\frac{\gamma}{1 - \gamma \tilde{\varepsilon}_{(b,b)}}\right)^M}{1 - \gamma} (V^G - u(b, b)). \]

It follows that \(\lim_{(\gamma, \varepsilon) \to (1,0)} (1 - \tilde{\varepsilon}_{(b,b)} - \varepsilon_{(a',a),(a,a)} - \varepsilon_{(a',a),(b,b)} \gamma (V^G - V^{B_0})/(1 - \gamma) = \lim_{(\gamma, \varepsilon) \to (1,0)} (1 - \tilde{\varepsilon}_{(b,b)} - \varepsilon_{(a',a),(a,a)} - \varepsilon_{(a',a),(b,b)} \gamma (V^G - V^{B_0})/(1 - \gamma) = M(u(a, a) - u(b, b)).\) Since \(M(u(a, a) - u(b, b)) > \max_{a'} \{\max\{u(a', a) - u(a, a), u(a', b) - u(b, b)\}\},\) good-standing players’ incentives are satisfied for \((\gamma, \varepsilon)\) sufficiently close to \((1,0).\)
A.2.2 Proof of Theorem 2(ii)

We show that $a$ can be limit-supported by the threaded grim trigger strategies discussed in Section 3, and then show that the constructed equilibria are coordination-proof.

A.2.2.1 Proof that $a$ is Limit-Supported by Strict Equilibria

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

$$\frac{\gamma}{1 - \gamma} > \max \left\{ \max_{(x_1, x_2)} \frac{u(x_1, x_2) - u(a, a)}{u(a, a) - u(b, b)}, \max_{(x_1, x_2)} \frac{u(x_1, x_2) - u(b, b)}{u(a, a) - u(b, b)} \right\}$$

(4)

for all $\gamma \in [\underline{\gamma}, \bar{\gamma}]$. Consider the grim-trigger strategy described in Section 3, and let $\mu^G$ denote the share of good-standing players in a steady state. We will show that for all $\delta > 0$, there is an $\varepsilon > 0$ such that, whenever $\gamma \in [\underline{\gamma}, \bar{\gamma}]$ and $\varepsilon(x_1, x_2), (x'_1, x'_2) < \varepsilon$ for all $(x_1, x_2), (x'_1, x'_2) \in A^2$, this strategy induces strict equilibria satisfying $\mu^G > 1 - \delta$.

Thus, this strategy can be combined with the threading technique described in the text to limit-support $a$ as $(\gamma, \varepsilon) \to (1, 0)$.

First we establish that for all $\delta > 0$, there is an $\varepsilon > 0$ such that, whenever $\gamma \in [\underline{\gamma}, \bar{\gamma}]$ and $\varepsilon(x_1, x_2), (x'_1, x'_2) < \varepsilon$ for all $(x_1, x_2), (x'_1, x'_2) \in A^2$, the steady states induced by this strategy satisfy $\mu^G > 1 - \delta$. Note that the inflow into good standing is simply $1 - \gamma$, 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 $\gamma(\varepsilon(a,a)\mu^G + \varepsilon(b,b)(1 - \mu^G))\mu^G$, the share of good-standing players whose outcome is recorded as something other than $(a, a)$ or $(b, b)$ 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

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

The claim then follows since $\lim_{\varepsilon \to 0} \inf_{\gamma \in [\underline{\gamma}, \bar{\gamma}]}(1 - \gamma)/(1 - \gamma + \gamma \max\{\varepsilon(a,a), \varepsilon(b,b)\}) = 1$.

We establish that for all $\delta > 0$, there is an $\varepsilon > 0$ such that, whenever $\gamma \in [\underline{\gamma}, \bar{\gamma}]$ and $\varepsilon(x_1, x_2), (x'_1, x'_2) < \varepsilon$ for all $(x_1, x_2), (x'_1, x'_2) \in A^2$, the incentives of good-standing
players are satisfied. (Since \((b, b)\) is a strict static equilibrium, the incentives of the bad-standing players are always satisfied.) The value function of good-standing players, \(V^G\), equals the average flow payoff in the population in a given period (since newborn players are in good standing), so 
\[
V^G = \mu^G(u(a, a) + (1 - \mu^G)u(b, b)) + (1 - \mu^G)u(b, b).
\]
In contrast, the value function of bad-standing players, \(V^B\), equals the expected flow payoff of bad-standing players, so 
\[
V^B = u(b, b).
\]

When facing an opponent playing \(a\), the expected payoff of a good-standing player from playing \(a\) is 
\[
(1 - \gamma)u(a, a) + \gamma(1 - \bar{\epsilon}_{(a, a)})V^G + \bar{\epsilon}_{(a, a)}V^B
\]
while their expected payoff from playing \(x \neq a\) is 
\[
(1 - \gamma)u(x, a) + \gamma(\epsilon_{(x, a), (a, a)} + \epsilon_{(x, a), (b, b)})V^G + (1 - \epsilon_{(x, a), (a, a)} - \epsilon_{(x, a), (b, b)})V^B.
\]
Thus, a good-standing player strictly prefers to play \(a\) rather than any \(x \neq a\) precisely when
\[
\frac{\gamma}{1 - \gamma} > \max_{x \neq a} \frac{u(x, a) - u(a, a)}{(1 - \bar{\epsilon}_{(a, a)} - \bar{\epsilon}_{(x, a), (a, a)} - \epsilon_{(x, a), (b, b)})\mu^G(u(a, a) - u(b, b)).
\]

As \(\epsilon \to 0\), the right-hand side of this inequality converges to \(\max_{x \neq a} (u(x, a) - u(a, a))/(u(a, a) - u(c, b))\), uniformly over \(\gamma \in [\underline{\gamma}, \bar{\gamma}]\). By the inequality in (4), we conclude that a good-standing player strictly prefers to match \(a\) with \(a\) instead of playing some \(x \neq a\) for sufficiently small noise when \(\gamma \in [\underline{\gamma}, \bar{\gamma}]\).

When facing an opponent playing \(b\), the expected payoff of a good-standing player from playing \(b\) is 
\[
(1 - \gamma)u(b, b) + \gamma(1 - \bar{\epsilon}_{(b, b)})V^G + \bar{\epsilon}_{(b, b)}V^B
\]
while their expected payoff from playing \(x \neq b\) is 
\[
(1 - \gamma)u(x, b) + \gamma(\epsilon_{(x, b), (a, a)} + \epsilon_{(x, b), (b, b)})V^G + (1 - \epsilon_{(x, b), (a, a)} - \epsilon_{(x, b), (b, b)})V^B.
\]
Thus a good-standing player strictly prefers to play \(a\) rather than any \(x \neq b\) precisely when
\[
\frac{\gamma}{1 - \gamma} > \max_{x \neq b} \frac{u(x, b) - u(b, b)}{(1 - \bar{\epsilon}_{(b, b)} - \bar{\epsilon}_{(x, b), (a, a)} - \epsilon_{(x, b), (b, b)})\mu^G(u(a, a) - u(b, b)).
\]

As \(\epsilon \to 0\), the right-hand side of this inequality converges to \(\max_{x \neq b} (u(x, b) - u(b, b))/(u(a, a) - u(c, b))\), uniformly over \(\gamma \in [\underline{\gamma}, \bar{\gamma}]\). By the inequality in (4), we conclude that a good-standing player strictly prefers to match \(b\) with \(b\) instead of playing some \(x \neq b\) for
sufficiently small noise when $\gamma \in [\gamma, \overline{\gamma}]$.

A.2.2.2 Proof of Coordination-Proofness

We first argue that in every match between bad-standing players, there is no Nash equilibrium in the augmented game that Pareto-dominates $(b, b)$. Note that the outcome of the current match does not affect a bad-standing player’s continuation value. Thus, any Nash equilibrium in the augmented game between two bad-standing players must also be a static equilibrium in the stage game. Since there is no static equilibrium that Pareto-dominates $(b, b)$, it follows that two bad-standing players playing $(b, b)$ is coordination-proof.

Now we show that in any match involving a good-standing player, there is no Nash equilibrium in the augmented game that Pareto-dominates the action profile the players are supposed to play. A very similar argument to that showing that a good-standing player strictly prefers to play $a$ against $a$ for sufficiently small noise when $\gamma \in [\gamma, \overline{\gamma}]$ shows that no good-standing player would ever prefer an action profile other than $(a, a)$ or $(b, b)$ be played in one of their matches. Thus, in any match involving a good-standing player, we need only consider whether $(a, a)$ or $(b, b)$ are Nash equilibria in the augmented game and whether one of these profiles Pareto-dominates the other. When two good-standing players match, both $(a, a)$ and $(b, b)$ are Nash equilibria in the augmented game, but $(b, b)$ does not Pareto-dominate $(a, a)$. Indeed, if $(b, b)$ did Pareto-dominate $(a, a)$, this would imply that the value functions for these good-standing players would be no higher than $u(b, b)$, which is not possible given that $u(a, a) > u(b, b)$. Thus, the prescribed play between two good-standing players is coordination-proof. Moreover, in any match involving a bad-standing player, all Nash equilibria in the augmented game require the bad-standing player to play a static best-response to the action of their opponent. Because $u(a, a) > u(b, b)$ and $(b, b)$ is not Pareto-dominated by any static equilibrium, $(a, a)$ is not a static equilibrium, so $(b, b)$ is coordination-proof when a good-standing player matches a bad-standing player.
A.3 Proof of Theorem 5

Section A.3.1 derives the incentive constraints that must be satisfied in any strict equilibrium with noisy first-order records, and Section A.3.2 proves Theorem 5(i) (necessary conditions for cooperation). The main step is proving Lemma 9, which shows that \( \mu_P + \mu_S(l - g) > g \) in any strict, coordination-proof equilibrium with \( \mu_C > 0 \). Section A.3.3 proves Theorem 5(ii) (sufficient conditions for cooperation). This part of the proof is split into three parts: Section A.3.3.1 shows that threaded grim trigger strategies can limit-support cooperation when \( g < 1 \); Section A.3.3.2 shows that threaded “defector → preciprocator → supercooperator → defector” strategies can limit-support cooperation when \( l > g + g^2 \); and Section A.3.3.3 shows that each class of equilibria is coordination-proof.

A.3.1 Incentive Constraints with Noisy Records

Throughout, \((C|C)_r\) denotes the condition that \( C \) is the best response to \( C \) for a player with record \( r \), \((C|D)_r\) denotes the condition that \( C \) is the best response to \( D \), and \((D|D)_r\) the condition that \( D \) is the best response to \( D \).

Let \( V^C_r \) denote the expected continuation payoff when a recording of \( C \) is fed into the record system for a record \( r \) player. That is, \( V^C_r = E_{r' \sim q_C(r)}[V_{r'}] \), where \( E_{r' \sim q_C(r)} \) indicates the expectation when \( r' \) is distributed according to \( q_C(r) \). Similarly, let \( V^D_r = E_{r' \sim q_D(r)}[V_{r'}] \) denote the expected continuation payoff when a recording of \( D \) is fed into the record system. Let \( \pi_r \) denote the expected flow payoff to a record \( r \) player under the equilibrium strategy, and let \( p^D_r \) denote the probability that a recording of \( D \) will be fed into the record system for a record \( k \) player. Note that \( p^D_r > 0 \) for all \( r \) since \( \varepsilon_C(r) > 0 \) and \( \varepsilon_D(r) < 1 \).

Given a noisy record system and an equilibrium, define the normalized reward for playing \( C \) rather than \( D \) for a record \( r \) player by

\[
W_r := \frac{1 - \varepsilon_C(r) - \varepsilon_D(r)}{p^D_r} \left( \pi_r - V_r + \frac{\gamma}{1 - \gamma} (V^C_r - V_r) \right).
\]
Lemma 6. For any noisy record system,

- The \((C|C)_r\) constraint is \(W_r > g\).
- The \((D|C)_r\) constraint is \(W_r < g\).
- The \((C|D)_r\) constraint is \(W_r > l\).
- The \((D|D)_r\) constraint is \(W_r < l\).

Proof. Consider a player with record \(r\). We derive the \((C|C)_r\) constraint; the other constraints can be similarly derived. When a record \(r\) player plays \(C\), their expected continuation payoff is \((1 - \varepsilon_C(r))V_r^C + \varepsilon_C(r)V_r^D\), since a recording of \(C\) is fed into the record system with probability \(1 - \varepsilon_C(r)\) and a recording of \(D\) is fed into the record system with probability \(\varepsilon_C(r)\). Similarly, when the player plays \(D\), their expected continuation payoff is \(\varepsilon_D(r)V_r^C + (1 - \varepsilon_D(r))V_r^D\). Thus, the \((C|C)_r\) constraint is \(1 - \gamma + \gamma(1 - \varepsilon_C(r))V_r^C + \gamma\varepsilon_C(r)V_r^D > (1 - \gamma)(1 + g) + \gamma\varepsilon_D(r)V_r^C + (1 - \varepsilon_D(r))V_r^D\), which is equivalent to

\[
(1 - \varepsilon_C(r) - \varepsilon_D(r))\frac{\gamma}{1 - \gamma}(V_r^C - V_r^D) > g.
\]

Note that \(V_r = (1 - \gamma)\pi_r + \gamma(1 - p_r^D)V_r^C + \gamma p_r^D V_r^D\). Manipulating this gives \(V_r^C - V_r^D = ((1 - \gamma)\pi_r - V_r + \gamma V_r^C)/(\gamma p_r^D)\). Substituting this into the above inequality gives the desired form of the \((C|C)_r\) constraint. 

The strategies we use to prove part (ii) of the theorem depend on a player’s record only through their age and their “score”, which is the number of times they have been recorded as playing \(D\). For such scoring strategies, we slightly abuse notation in writing \(V_k\) for the continuation payoff of a player with score \(k\).\(^{24}\) The incentive constraints take a simpler form with such strategies: For all \(k\) we have \(\varepsilon_C(k) = \varepsilon_C\),

\(^{24}\)Recall that \(V_r\) is defined as the continuation value of a player with record \(r\). Under scoring strategies, two players with different records that share the same score have the same continuation value, so we can index \(V\) by \(k\) rather than \(r\).
\( \varepsilon_D(k) = \varepsilon_D, \ V^C_k = V_k, \) and \( V^D_k = V_{k+1}. \) The normalized reward thus simplifies to

\[
W_k = \frac{1 - \varepsilon_C - \varepsilon_D}{p_k^D} (\pi_k - V_k).
\]

**Lemma 7.** For scoring strategies, Lemma 6 holds with \( W_k = (1 - \varepsilon_C - \varepsilon_D)(\pi_k - V_k)/p_k^D. \)

### A.3.2 Proof of Theorem 5(i)

Theorem 5(i) follows from the following two lemmas.

**Lemma 8.** For any first-order record system, in any strict equilibrium, \( \mu^S < 1/(1 + g). \)

**Lemma 9.** For any noisy first-order record system, in any strict, coordination-proof equilibrium with \( \mu^C > 0, \mu^P + \mu^S (l - g) > g. \)

Lemma 8 says that there cannot be too many supercooperators. It holds because new players with record 0 have the option of always playing \( D, \) so in any strict equilibrium with \( \mu^C > 0, \) it must be that \( \mu^S (1 + g) < V_0 \leq 1, \) which gives \( \mu^S < 1/(1 + g) \)

Conversely, Lemma 9 implies that cooperation requires a positive share of supercooperators when \( g \geq 1, \) and moreover that the required share grows when \( g \) and \( l \) are increased by the same amount. It is proved in the next subsection.

Theorem 5(i) follows from Lemmas 8 and 9 since, by Lemma 3, it is impossible to satisfy both \( \mu^S < 1/(1 + g) \) and \( \mu^P + \mu^S (l - g) > g \) when \( g \geq 1 \) and \( l \leq g + g^2. \)

#### A.3.2.1 Necessary Conditions for Cooperation and Proof of Lemma 9

Let \( \bar{V} = \sup, V_r \) and let \( \{r_n\}_{n \in \mathbb{N}} \) be a sequence of records such that \( \lim_{n \to \infty} V_{r_n} = \bar{V}. \) Note that \( \bar{V} < \infty \) and, since \( V_0 \) (the expected lifetime payoff of a newborn player) equals \( \mu^P \mu^C + \mu^S (\mu^C - \mu^D l) + \mu^D \mu^S (1 + g) \) (the average flow payoff in the population), we have \( \bar{V} \geq V_0 = \mu^P \mu^C + \mu^S (\mu^C - \mu^D l) + \mu^D \mu^S (1 + g). \)

**Lemma 10.** If \( \mu^C > 0, \) there is no sequence of defector records \( \{r_n\}_{n \in \mathbb{N}} \) such that \( \lim_{n \to \infty} V_{r_n} = \bar{V}. \)
Proof. Suppose otherwise. Since \( V_r = (1 - \gamma)\pi_r + \gamma(1 - p^D_r)\mu^C_r + \gamma p^D_r\mu^D_r \) and \( \pi_{r_n} = \mu^S(1 + g) \) for all \( r_n \), we have \( V_{r_n} = (1 - \gamma)\mu^S(1 + g) + \gamma(1 - p^D_{r_n})\mu^C_{r_n} + \gamma p^D_{r_n}\mu^D_{r_n} \) for all \( r_n \). This implies

\[
V_{r_n} \leq \mu^S(1 + g) + \frac{\gamma}{1 - \gamma}(1 - p^D_{r_n})\max\{V^C_{r_n} - V_{r_n}, 0\} + \frac{\gamma}{1 - \gamma}p^D_{r_n}\max\{V^D_{r_n} - V_{r_n}, 0\}.
\]

Since \( \lim_{n \to \infty} V_{r_n} = V \), \( \lim_{n \to \infty} \max\{V^C_{r_n} - V_{r_n}, 0\} = \lim_{n \to \infty} \max\{V^D_{r_n} - V_{r_n}, 0\} = 0 \). It further follows that \( V = \lim_{n \to \infty} V_{r_n} \leq \mu^S(1 + g) \), so \( V_r \leq \mu^S(1 + g) \) for all \( r \). However, note that every player can secure an expected flow payoff of \( \mu^S(1 + g) \) every period by always defecting, so it must be that \( V_r \geq \mu^S(1 + g) \) for all \( r \). It follows that \( V_r = \mu^S(1 + g) \) for all \( r \), and since the value function is constant across records, every record must be a defector record, so \( \mu^C = 0 \). \( \blacksquare \)

**Lemma 11.** If \( \mu^C > 0 \), there is some record \( r' \) that is a preciprocator or a supercooperator and satisfies

\[
V_{r'} - \frac{\gamma}{1 - \gamma}(V^C_{r'} - V_{r'}) \geq \mu^P \mu^C + \mu^S(\mu^C - \mu^D l) + \mu^D \mu^S(1 + g).
\]

Proof. First, consider the case where \( V = \mu^P \mu^C + \mu^S(\mu^C - \mu^D l) + \mu^D \mu^S(1 + g) \). Then there must be some record \( r' \) such that \( V_{r'} = \mu^P \mu^C + \mu^S(\mu^C - \mu^D l) + \mu^D \mu^S(1 + g) \). By Lemma 10, such a \( r' \) cannot be a defector record and so must be either a preciprocator or a supercooperator. Additionally, \( V^C_{r'} \leq V \), so \( V_{r'} - (\gamma/(1 - \gamma))(V^C_{r'} - V_{r'}) \geq \mu^P \mu^C + \mu^S(\mu^C - \mu^D l) + \mu^D \mu^S(1 + g) \).

Now consider the case where \( V > \mu^P \mu^C + \mu^S(\mu^C - \mu^D l) + \mu^D \mu^S(1 + g) \). For any sequence of records \( \{r_n\}_{n \in \mathbb{N}} \) such that \( \lim_{n \to \infty} V_{r_n} = V \), \( \lim_{n \to \infty} \max\{V^C_{r_n} - V_{r_n}, 0\} = 0 \), so there is some sufficiently high \( n \) such that \( V_{r_n} - (\gamma/(1 - \gamma))(V^C_{r_n} - V_{r_n}) \geq \mu^P \mu^C + \mu^S(\mu^C - \mu^D l) + \mu^D \mu^S(1 + g) \). Additionally, by Lemma 10, for sufficiently high \( n \), the record \( r_n \) must be either a preciprocator or a supercooperator. \( \blacksquare \)

**Proof of Lemma 9.** First, take the case where \( r' \) is a preciprocator. Then by Lemma
6, we must have

\[
\frac{1 - \varepsilon_C(r') - \varepsilon_D(r')}{{p_{r'}}^D} \left( \pi_{r'} - V_{r'} + \frac{\gamma}{1 - \gamma} (V_{r'}^C - V_{r'}) \right) > g.
\]

When \( \pi_{r'} = \mu^C \) and \( V_{r'} - \gamma (V_{r'}^C - V_{r'})/(1 - \gamma) \geq \mu^P \mu^C + \mu^S (\mu^C - \mu^D l) + \mu^D \mu^S (1 + g) \), this implies

\[
\frac{(1 - \varepsilon_C(r') - \varepsilon_D(r')){\mu^D}}{{p_{r'}}^D} (\mu^P + \mu^S (l - g)) > g.
\]

Note that \( {p_{r'}}^D \geq (1 - \varepsilon_D(r')){\mu^D} \) since a preciprocator plays \( D \) whenever they are matched with a defector and this leads to a recording of \( D \) being fed into the record system with probability \( 1 - \varepsilon_D(r') \). This gives \( (1 - \varepsilon_C(r') - \varepsilon_D(r')){\mu^D} / {p_{r'}}^D < 1 \), so \( \mu^P + \mu^S (l - g) > g \) must hold.

Now take the case where \( r' \) is a supercooperator. By Lemma 6, \( \pi_{r'} - V_{r'} + (\gamma/(1 - \gamma)) (V_{r'}^C - V_{r'}) > 0 \). When \( \pi_{r'} = \mu^C - \mu^D l \) and \( V_{r'} - (\gamma/(1 - \gamma)) (V_{r'}^C - V_{r'}) \geq \mu^P \mu^C + \mu^S (\mu^C - \mu^D l) + \mu^D \mu^S (1 + g) \), this implies that

\[
\mu^C - \mu^D l - (\mu^P \mu^C + \mu^S (\mu^C - \mu^D l) + \mu^D \mu^S (1 + g)) = \mu^D (\mu^P + \mu^S (l - g) - l) > 0.
\]

This requires \( \mu^P + \mu^S (l - g) > l \), which implies \( \mu^P + \mu^S (l - g) > g \), since \( l > g \).

\section*{A.3.3 Proof of Theorem 5(ii)}

\subsection*{A.3.3.1 Limit-Supporting \( C \) when \( g < 1 \)}

Let \( 0 < \gamma < \overline{\gamma} < 1/2 \) be such that

\[
g < \frac{\gamma}{1 - \gamma} < l
\]

for all \( \gamma \in [\gamma, \overline{\gamma}] \). Consider the grim trigger strategy, and let \( \mu^C(\gamma, \varepsilon) \) denote the steady state share of cooperators, i.e. those players who have not been recorded as playing \( D \), for parameters \( (\gamma, \varepsilon) \). (As we will see, there is a unique steady state when noise
is sufficiently small.) We will show that for all \( \delta > 0 \), there is an \( \varepsilon > 0 \) such that, whenever \( \gamma \in [\underline{\gamma}, \bar{\gamma}] \) and \( \varepsilon_C, \varepsilon_D < \varepsilon \), this strategy induces strict equilibria satisfying \( \mu^C(\gamma, \varepsilon) > 1 - \delta \). Thus, this strategy can be combined with threading to limit-support \( C \) as \( (\gamma, \varepsilon) \to (1, 0) \).

First, we establish that for all \( \delta > 0 \), there is an \( \varepsilon > 0 \) such that whenever \( \gamma \in [\underline{\gamma}, \bar{\gamma}] \) and \( \varepsilon_C, \varepsilon_D < \varepsilon \), the steady states induced by this strategy satisfies \( \mu^C > 1 - \delta \). Note that the inflow into cooperator status is \( 1 - \gamma \), the share of newborn players. The outflow from cooperator status is the sum of \( (1 - \gamma)\mu^C \), the share of cooperators who die in a given period, and \( \gamma(\varepsilon_C\mu^C + (1 - \varepsilon_D)(1 - \mu^C))\mu^C \), the share of cooperators who are recorded as playing \( D \) in a given period. In a steady state, these inflows and outflows must be equal, so

\[
1 - \gamma = (1 - \gamma + \gamma(\varepsilon_C\mu^C + (1 - \varepsilon_D)(1 - \mu^C))\mu^C).
\]

This expression has a unique solution \( \mu^C \in [0, 1] \) when \( \varepsilon_C \) and \( \varepsilon_D \) are sufficiently small, given by

\[
\mu^C(\gamma, \varepsilon) = \frac{1 - \gamma\varepsilon_D - \sqrt{(1 - \gamma\varepsilon_D)^2 - 4\gamma(1 - \varepsilon_C - \varepsilon_D)(1 - \gamma)}}{2\gamma(1 - \varepsilon_C - \varepsilon_D)}.
\]

Note that \( \mu^C(\gamma, \varepsilon) \) is continuous for \( \gamma \in [\underline{\gamma}, \bar{\gamma}] \) and sufficiently small \( \varepsilon_C, \varepsilon_D \), and \( \mu^C(\gamma, 0) = 1 \) for all \( \gamma \leq 1/2 \). It follows that there is an \( \varepsilon > 0 \) such that \( \mu^C(\gamma, \varepsilon) > 1 - \delta \) for all \( \gamma \in [\underline{\gamma}, \bar{\gamma}] \) and \( \varepsilon_C, \varepsilon_D < \varepsilon \).

Now we establish that for all \( \delta > 0 \), there is an \( \varepsilon > 0 \) such that whenever \( \gamma \in [\underline{\gamma}, \bar{\gamma}] \) and \( \varepsilon_C, \varepsilon_D < \varepsilon \), the incentives of preciprocators are satisfied. (The incentives of defectors are clearly satisfied.) We will use the facts that the value function of preciprocators, \( V^C \), equals the average flow payoff in the population in a given period, \( (\mu^C(\gamma, \varepsilon))^2 \), and that the value function of defectors is \( V^D = 0 \).

When facing an opponent playing \( C \), the expected payoff for a preciprocator from playing \( C \) is \( 1 - \gamma + \gamma(1 - \varepsilon_C)(\mu^C(\gamma, \varepsilon))^2 \) while their expected payoff from playing \( D \) is
\[(1 - \gamma)(1 + g) + \gamma \varepsilon_D(\mu^C(\gamma, \varepsilon))^2.\] Thus, a preciprocator strictly prefers to play \(C\) against an opponent playing \(C\) if and only if \[1 - \gamma + \gamma(1 - \varepsilon_C)(\mu^C(\gamma, \varepsilon))^2 > (1 - \gamma)(1 + g) + \gamma \varepsilon_D(\mu^C(\gamma, \varepsilon))^2,\] which simplifies to

\[
\frac{\gamma}{1 - \gamma} > \frac{g}{(1 - \varepsilon_C - \varepsilon_D)(\mu^C(\gamma, \varepsilon))^2}.
\]

When facing an opponent playing \(D\), the expected payoff to \(C\) for a preciprocator is \[-(1 - \gamma)l + \gamma(1 - \varepsilon_C)(\mu^C(\gamma, \varepsilon))^2\] while their expected payoff from playing \(D\) is \[\gamma \varepsilon_D(\mu^C(\gamma, \varepsilon))^2.\] Thus, a preciprocator strictly prefers to play \(D\) against an opponent playing \(D\) if and only if \[-(1 - \gamma)l + \gamma(1 - \varepsilon_C)(\mu^C(\gamma, \varepsilon))^2 \leq \gamma \varepsilon_D(\mu^C(\gamma, \varepsilon))^2,\] which simplifies to

\[
\frac{\gamma}{1 - \gamma} < \frac{l}{(1 - \varepsilon_C - \varepsilon_D)(\mu^C(\gamma, \varepsilon))^2}.
\]

Combining these incentive conditions shows that all the incentives of a preciprocator are satisfied if and only if

\[
\frac{g}{(1 - \varepsilon_C - \varepsilon_D)(\mu^C(\gamma, \varepsilon))^2} \leq \frac{\gamma}{1 - \gamma} \leq \frac{l}{(1 - \varepsilon_C - \varepsilon_D)(\mu^C(\gamma, \varepsilon))^2}.
\]

As \(\varepsilon \to 0\), the left-most expression and right-most expression in this inequality converge to \(g\) and \(l\), respectively, uniformly over \(\gamma \in [\underline{\gamma}, \overline{\gamma}]\). By inequality (5), we conclude that the incentives of a preciprocator are satisfied for sufficiently small noise when \(\gamma \in [\underline{\gamma}, \overline{\gamma}]\).

**A.3.3.2 Limit-Supporting \(C\) when \(l > g + g^2\)**

We use the class of strategies of the form \("D_J P_K S_1 D_\infty,"\) where \(J, K \in \mathbb{N}\). These strategies specify that a player is a defector until they have been recorded as playing \(D\) \(J\) times. Subsequently, the player is a preciprocator until they have been recorded as playing \(D\) \(K\) more times, and then a supercooperator until they are recorded as playing \(D\) once more, after which they permanently become a defector. Throughout, we let \(\mu^{D_1}\) be the share of players who have been recorded as playing \(D\) fewer than \(J\) times (and are thus defectors), \(\mu^P\) be the share of preciprocators (those with score
$J \leq k < J + K$, $\mu^S$ be the share of supercooperators (those with score $k = J + K$), and $\mu^{D_2}$ be the share of defectors with a score $k > J + K$. We also let $\mu^C = \mu^P + \mu^S$ be the total share of cooperators and $\mu^D = \mu^{D_1} + \mu^{D_2} = 1 - \mu^C$ be the total share of defectors. We will show that for all $\delta > 0$, there are $0 < \gamma < \bar{\gamma} < 1$ and $\bar{\varepsilon} > 0$ such that when $\gamma \in [\gamma, \bar{\gamma}]$ and $\varepsilon_C, \varepsilon_D < \bar{\varepsilon}$, this strategy class gives equilibria satisfying $\mu^C > 1 - \delta$. Thus, these strategies can be combined with threading to limit-support $C$ as $(\gamma, \varepsilon) \to (1, 0)$.

The following lemma characterizes precisely which population shares and parameters are consistent with an equilibrium using a $D_J P_K S_1 D_\infty$ strategy. The statement of the lemma involves the functions $\alpha : (0, 1) \times (0, 1) \to (0, 1)$ and $\beta : (0, 1) \times (0, 1) \times [0, 1] \to (0, 1)$, defined by

\[
\alpha(\gamma, \psi) = \frac{\gamma \psi}{1 - \gamma + \gamma \psi},
\]

\[
\beta(\gamma, \varepsilon, \mu^D) = \frac{\gamma (\varepsilon_C + (1 - \varepsilon_C - \varepsilon_D) \mu^D)}{1 - \gamma + \gamma (\varepsilon_C + (1 - \varepsilon_C - \varepsilon_D) \mu^D)}.
\]

**Lemma 12.** There is a $D_J P_K S_1 D_\infty$ equilibrium with shares $\mu^{D_1}$, $\mu^P$, $\mu^S$, and $\mu^{D_2}$ 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:**

   \[
   (C|C)_J : \frac{(1 - \varepsilon_C - \varepsilon_D) \mu^D}{\varepsilon_C + (1 - \varepsilon_C - \varepsilon_D) \mu^D} \left( \frac{\mu^S}{1 - \mu^{D_1} l} + \frac{\mu^{D_2}}{\mu^{D_1} (1 - \mu^{D_1})} (\mu^P - \mu^S g) \right) > g,
   \]

   \[
   (D|D)_{J+K-1} : \frac{\gamma(1 - \varepsilon_C - \varepsilon_D)(1 - \alpha(\gamma, \varepsilon_C) \mu^D l + \alpha(\gamma, \varepsilon_C) (\mu^P - \mu^S g))}{1 - \gamma + \gamma (\varepsilon_C + (1 - \varepsilon_C - \varepsilon_D) \mu^D)} < l,
   \]

   \[
   (C|D)_{J+K} \ (\text{if } \mu^S > 0) : \frac{\gamma(1 - \varepsilon_C - \varepsilon_D)}{1 - \gamma + \gamma \varepsilon_C} (\mu^P - \mu^S g - \mu^D l) > l.
   \]
The proof of Lemma 12 is in OA.3.1. The feasibility constraints come from calculating the relevant steady-state shares for the strategy $D_J P_K S_I D_{\infty}$. The $(C|C)_J$ incentive constraint comes from solving $V_J$ and using Lemma 7. The $(C|D)_{J+K}$ and $(D|D)_{J+K-1}$ constraints are derived by relating the value functions of adjacent records.

Since $l > g + g^2$, it can be shown that, for all sufficiently small $\delta > 0$, there are $\bar{\mu}^P, \bar{\mu}^S > 0$ satisfying $\bar{\mu}^P + \bar{\mu}^S = 1 - \delta$, $\bar{\mu}^S > (g/l)(1 - \delta)$, and $\bar{\mu}^P - \bar{\mu}^S g - \delta l > 0$. Fix such a $\delta$ and the corresponding $\bar{\mu}^P, \bar{\mu}^S$. There is some sufficiently small $\eta \in (0, \delta/2)$ such that the above inequalities hold when $\bar{\mu}^P, \bar{\mu}^S$, and $\delta$ are respectively replaced with any $\mu^P, \mu^S$, and $\hat{\delta}$ satisfying $|\mu^P - \bar{\mu}^P|, |\mu^S - \bar{\mu}^S| \leq \eta$ and $|\hat{\delta} - \delta| \leq 2\eta$.

The following lemma, whose proof is in OA.3.2, shows that there is an interval of $\gamma$ such that $J$ and $K$ can be tailored to obtain shares $\mu^{D_1}, \mu^P$, and $\mu^S$ within $\eta$ of $\delta$, $\bar{\mu}^P$, and $\bar{\mu}^S$, respectively, when noise is sufficiently small. (Consequently, the share $\mu^D$ must be within $2\eta$ of $\delta$.) Moreover, the $\gamma$ interval can be taken so that the incentives of supercooperators are satisfied.

**Lemma 13.** There are $0 < \gamma < \bar{\gamma} < 1$ and $\bar{\varepsilon} > 0$ such that, for all $\gamma \in [\gamma, \bar{\gamma}]$ and $\varepsilon_C, \varepsilon_D < \bar{\varepsilon}$, there are steady states with shares satisfying $|\mu^D_1 - \delta|, |\mu^P - \bar{\mu}^P|, |\mu^S - \bar{\mu}^S| \leq \eta$, and are such that the $(C|D)_{J+K}$ constraint in Lemma 12 is satisfied.

The left-hand side of the $(C|C)_J$ constraint in Lemma 12 converges uniformly to $\mu^S/(1 - \mu^{D_1})l$ as $\varepsilon \to 0$ for all $\gamma \in [\gamma, \bar{\gamma}]$, $|\mu^{D_1} - \delta|, |\mu^P - \bar{\mu}^P|, |\mu^S - \bar{\mu}^S| \leq \eta$. Because $(\bar{\mu}^S - \eta)/(1 - \delta + \varepsilon l) > g$, this means that $\bar{\varepsilon}$ can be chosen to be sufficiently small such that all these steady states satisfy the $(C|C)_J$ constraint in Lemma 12 for all $\gamma \in [\gamma, \bar{\gamma}]$ and $\varepsilon_C, \varepsilon_D < \bar{\varepsilon}$. This is similarly true for the $(D|D)_{J+K-1}$ constraint in Lemma 12, because the left-hand side of the corresponding inequality converges uniformly to $\gamma \mu^D/(1 - \gamma + \gamma \mu^D)l < l$ as $\varepsilon \to 0$ for all $\gamma \in [\gamma, \bar{\gamma}]$, $|\mu^{D_1} - \delta|, |\mu^P - \bar{\mu}^P|, |\mu^S - \bar{\mu}^S| \leq \eta$.

Thus there are $0 < \gamma < \bar{\gamma} < 1$ and $\bar{\varepsilon} > 0$ such that equilibria with shares $\mu^P, \mu^S$ satisfying $|\mu^P - \bar{\mu}^P|, |\mu^S - \bar{\mu}^S| \leq \eta$ (and thus $\mu^C \geq 1 - 2\delta$) exist whenever $\gamma \in [\gamma, \bar{\gamma}]$ and $\varepsilon_C, \varepsilon_D < \bar{\varepsilon}$. 

49
A.3.3.3 Proof of Coordination-Proofness

We show that the grim trigger equilibria analyzed in A.3.3.1 and the $D_JP_KS_1D_\infty$ equilibria analyzed in A.3.3.2 are coordination-proof. In any such equilibrium, $(C, C)$ is played in every match where neither player has a defector record. By a similar argument to the proof of Lemma 2, the play in these matches is coordination-proof. Thus, we need only consider play in matches with a defector. Note that in equilibria generated by either grim trigger or $D_JP_KS_1D_\infty$ strategies, the expected continuation value of a defector is weakly higher from playing $D$ than from playing $C$. Since $D$ is strictly dominant in the stage game, it follows that $D$ is strictly dominant in the augmented game for any defector. Thus, the prescribed action profile $(D, D)$ in a match involving a preciprocator and a defector is the only equilibrium in the corresponding augmented game. Likewise, the prescribed action profile $(C, D)$ in a match involving a supercooperator and a defector is the only equilibrium in the corresponding augmented game. We conclude that play in all matches is coordination-proof.
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 Nash equilibrium.

Fix such a symmetric static equilibrium $\alpha^*$, and let $\sigma$ recommend $\alpha^*$ at every record pair $(r, r')$. Then $(\sigma, \mu)$ is an equilibrium for any steady state $\mu$. 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, $(\alpha, \alpha')$ is a (possibly mixed) augmented-game Nash equilibrium if and only if it is a Nash equilibrium of the stage game. Since $(\alpha^*, \alpha^*)$ is not Pareto-dominated by another static equilibrium, there is no augmented-game Nash equilibrium $(\alpha, \alpha')$ satisfying $(u(\alpha, \alpha'), u(\alpha', \alpha)) >$
(u(\alpha^*, \alpha^*), u(\alpha^*, \alpha^*))$, and hence there is no augmented-game Nash equilibrium $(\alpha, \alpha')$ 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, $(\sigma, \mu)$ is coordination-proof.

### OA.2 Proof of Theorem 3

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

1. 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.

2. 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 $0 < \gamma < \overline{\gamma} < 1$ be such that

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

for all $\gamma \in [\gamma, \overline{\gamma}]$. Consider the strategy described in Section 4, and let $\mu^G$ denote the share of good-standing players in a steady state. We will show that for all $\delta > 0$, there is an $\varepsilon > 0$ such that, whenever $\gamma \in [\gamma, \overline{\gamma}]$ and $\varepsilon_{x,x'}, < \varepsilon$ for all $x, x' \in A$, this strategy induces strict equilibria satisfying $\mu^G > 1 - \delta$. Thus, this strategy can be combined with threading to limit-support $a$ as $(\gamma, \varepsilon) \rightarrow (1, 0)$.

First, we establish that for all $\delta > 0$, there is an $\varepsilon > 0$ such that, whenever $\gamma \in [\gamma, \overline{\gamma}]$ and $\varepsilon_{x,x'}, < \varepsilon$ for all $x, x' \in A$, the steady states induced by this strategy satisfies $\mu^G > 1 - \delta$. 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 \( \lim_{\varepsilon \to 0} \inf_{\gamma \in [\gamma, \eta]} (1 - \gamma) / (1 - \gamma + \gamma \max\{\tilde{\varepsilon}_a, \tilde{\varepsilon}_b\}) = 1 \).

Now we establish that, for all \( \delta > 0 \), there is an \( \varepsilon > 0 \) such that, whenever \( \gamma \in [\gamma, \eta] \) and \( \varepsilon_{x,x'}, < \varepsilon \) for all \( x, x' \in A \), the incentives of good-standing players states are satisfied. (Since \( c \) is a strict best-response to \( b \) and \( (d, d) \) is a strict static equilibrium, the incentives of bad-standing players are always satisfied.) 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

\[ (1 - \gamma)(u(a, a) - u(b, b)) > \gamma(\tilde{\varepsilon}_a - \tilde{\varepsilon}_b)(V^G - V^B). \]

As \( \varepsilon \to 0 \), the right-hand side of this inequality converges to 0, uniformly over \( \gamma \in [\gamma, \eta] \). As \( u(a, a) > u(b, b) \), we conclude that a good-standing player strictly prefers to match \( a \) with \( a \) instead of playing \( b \) for sufficiently small noise when \( \gamma \in [\gamma, \eta] \). Moreover, the expected payoff of a good-standing player from playing action \( x \not\in \{a, b\} \) 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 \( x \not\in \{a, b\} \) precisely when

\[ \gamma > \max_{x \not\in \{a, b\}} \frac{u(x, a) - u(a, a)}{(1 - \varepsilon_a - \varepsilon_{x,a} - \varepsilon_{x,b})(\mu^G(u(a, a) - u(c, b)) + (1 - \mu^G)(u(b, c) - u(d, d)))}. \]

As \( \varepsilon \to 0 \), the right-hand side of this inequality converges to \( \max_{x \not\in \{a, b\}} (u(x, a) - u(a, a)) \).
\[ u(a, a)/(u(a, a) - u(c, b)), \text{ uniformly over } \gamma \in [\underline{\gamma}, \overline{\gamma}]. \] By inequality (6), we conclude that a good-standing player strictly prefers to match \( a \) with \( a \) instead of playing some \( x \not\in \{a, b\} \) for sufficiently small noise when \( \gamma \in [\underline{\gamma}, \overline{\gamma}] \).

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

\[
(1 - \gamma)(u(b, c) - u(a, c)) > \gamma(\tilde{\varepsilon}_b - \tilde{\varepsilon}_a)(V^G - V^B).
\]

As \( \varepsilon \to 0 \), the right-hand side of this inequality converges to 0, uniformly over \( \gamma \in [\underline{\gamma}, \overline{\gamma}] \). As \( u(b, c) > u(a, c) \), we conclude that a good-standing player strictly prefers to play \( b \) rather than \( a \) against an opponent playing \( c \) for sufficiently small noise when \( \gamma \in [\underline{\gamma}, \overline{\gamma}] \).

Moreover, the expected payoff of a good-standing player from playing action \( x \not\in \{a, b\} \) 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 \( x \not\in \{a, b\} \) precisely when

\[
\frac{\gamma}{1 - \gamma} > \max_{x \not\in \{a, b\}} \frac{u(x, c) - u(b, c)}{(1 - \tilde{\varepsilon}_b - \varepsilon_{x, a} - \varepsilon_{x, b})(\mu^G(u(a, a) - u(c, b)) + (1 - \mu^G)(u(b, c) - u(d, d))}.
\]

As \( \varepsilon \to 0 \), the right-hand side of this inequality converges to \( \max_{x \not\in \{a, b\}} (u(x, c) - u(b, c))/(u(a, a) - u(c, b)) \), uniformly over \( \gamma \in [\underline{\gamma}, \overline{\gamma}] \). By Inequality 6, we conclude that a good-standing player strictly prefers to play \( b \) rather than some \( x \not\in \{a, b\} \) against an opponent playing \( c \) for sufficiently small noise when \( \gamma \in [\underline{\gamma}, \overline{\gamma}] \).
OA.3 Proofs of Lemmas for Theorem 5(ii)

OA.3.1 Proof of Lemma 12

Lemma 12. There is a $D_jP_KS_1D_\infty$ equilibrium with shares $\mu^{D_1}$, $\mu^P$, $\mu^S$, and $\mu^{D_2}$ if and only if the following conditions hold:

1. Feasibility:
   \[ \mu^{D_1} = 1 - \alpha(\gamma, 1 - \epsilon_D)^J, \]
   \[ \mu^P = \alpha(\gamma, 1 - \epsilon_D)^J (1 - \beta(\gamma, \epsilon, \mu^K)) , \]
   \[ \mu^S = \alpha(\gamma, 1 - \epsilon_D)^J \beta(\gamma, \epsilon, \mu^K) (1 - \alpha(\gamma, \epsilon_C)), \]
   \[ \mu^{D_2} = \alpha(\gamma, 1 - \epsilon_D)^J \beta(\gamma, \epsilon, \mu^K) \alpha(\gamma, \epsilon_C). \]

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

We will derive the feasibility conditions and then derive the incentive conditions.

The other feasibility conditions of Lemma 12 follow from the following lemma.

Lemma OA 1. In a $D_jP_KS_1D_\infty$ steady state with total share of defectors $\mu^D$, 

\[
\mu_k = \begin{cases} 
\alpha(\gamma, 1 - \epsilon_D)^k(1 - \alpha(\gamma, 1 - \epsilon_D)) & \text{if } 0 \leq k \leq J - 1 \\
\alpha(\gamma, 1 - \epsilon_D)^J \beta(\gamma, \epsilon, \mu^K)^k(1 - \beta(\gamma, \epsilon, \mu^K)) & \text{if } J \leq k \leq J + K - 1 \\
\alpha(\gamma, 1 - \epsilon_D)^J \beta(\gamma, \epsilon, \mu^K)^K(1 - \alpha(\gamma, \epsilon_C)) & \text{if } k = J + K 
\end{cases}
\]
To see why Lemma OA 1 implies the feasibility conditions of Lemma 12, note that

\[
\mu_{D_1} = \sum_{k=0}^{J-1} \alpha(\gamma, 1 - \varepsilon_D)^k(1 - \alpha(\gamma, 1 - \varepsilon_D)) = 1 - \alpha(\gamma, 1 - \varepsilon_D)^J,
\]

\[
\mu_P = \sum_{k=J}^{J+K-1} \alpha(\gamma, 1 - \varepsilon_D)^k(1 - \beta(\gamma, \varepsilon, \mu^D)) = \alpha(\gamma, 1 - \varepsilon_D)^J(1 - \beta(\gamma, \varepsilon, \mu^D)^K),
\]

\[
\mu_S = \mu_{J+K} = \alpha(\gamma, 1 - \varepsilon_D)^J(1 - \alpha(\gamma, \varepsilon)),
\]

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

\[
\mu_J = \alpha(\gamma, 1 - \varepsilon_D)^{J-1}(1 - \alpha(\gamma, 1 - \varepsilon_D))\frac{\gamma(1 - \varepsilon_D)}{1 - \gamma + \gamma(\varepsilon_C + (1 - \varepsilon_C - \varepsilon_D)\mu^D)}
\]

\[
= \alpha(\gamma, 1 - \varepsilon_D)^J \frac{1 - \gamma}{1 - \gamma + \gamma(\varepsilon_C + (1 - \varepsilon_C - \varepsilon_D)\mu^D)}
\]

\[
= \alpha(\gamma, 1 - \varepsilon_D)^J(1 - \beta(\gamma, \varepsilon, \mu^D)).
\]
Additionally, for every $J < k < J+K$, both record $k$ and record $k-1$ are precipractors. 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 precipractor 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 $\mu_{J+K-1} = \alpha(\gamma, 1 - \varepsilon_D)^J \beta(\gamma, \varepsilon, \mu^D)^K (1 - \beta(\gamma, \varepsilon, \mu^D))$, we have

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

$$
= \alpha(\gamma, 1 - \varepsilon_D)^J \beta(\gamma, \varepsilon, \mu^D)^K \left(\frac{1 - \gamma}{1 - \gamma + \gamma \varepsilon_C}\right)
$$

$$
= \alpha(\gamma, 1 - \varepsilon_D)^J \beta(\gamma, \varepsilon, \mu^D)^K (1 - \alpha(\gamma, \varepsilon_C)).
$$

Now we establish the incentive conditions in Lemma 12. We first handle the incentives of the score $J$ precipractor to play $C$ against an opponent playing $C$. (When this incentive condition is satisfied, all other precipractors 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

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

Since the flow payoff to a precipractor is $\mu^C$, Lemma 7 along with the fact that $p_k^D = \varepsilon_C + (1 - \varepsilon_C - \varepsilon_D)\mu^D$ for any precipractor implies that a score $J$ precipractor
plays $C$ against $C$ iff
\[
\frac{1 - \varepsilon_C - \varepsilon_D}{\varepsilon_C + (1 - \varepsilon_C - \varepsilon_D)} \left( \mu_C - \frac{\mu^P}{1 - \mu^D_1} \mu_C - \frac{\mu^S}{1 - \mu^D_1} (\mu_C - \mu^D l) - \frac{\mu^D_2}{1 - \mu^D_1} \mu^S (1 + g) \right) > g.
\]

Since
\[
\mu_C - \frac{\mu^P}{1 - \mu^D_1} \mu_C - \frac{\mu^S}{1 - \mu^D_1} (\mu_C - \mu^D l) - \frac{\mu^D_2}{1 - \mu^D_1} \mu^S (1 + g) = \mu^D \left( \frac{\mu^S}{1 - \mu^D_1} l + \frac{\mu^D_2}{\mu^D_1(1 - \mu^D_1)} (\mu^P - \mu^S g) \right),
\]

it follows that the $(C|C)_J$ constraint is equivalent to
\[
\frac{(1 - \varepsilon_C - \varepsilon_D) \mu^P}{\varepsilon_C + (1 - \varepsilon_C - \varepsilon_D) \mu^D} \left( \frac{\mu^S}{1 - \mu^D_1} l + \frac{\mu^D_2}{\mu^D_1(1 - \mu^D_1)} (\mu^P - \mu^S g) \right) > g.
\]

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
\[
V_{J+K} = (1 - \alpha(\gamma, \varepsilon_C))(\mu_C - \mu^D l) + \alpha(\gamma, \varepsilon_C) \mu^S (1 + g).
\]  \hspace{1cm} \text{(OA 1)}

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 1 gives

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

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 $0 < \gamma < \overline{\gamma} < 1$ and $\varepsilon > 0$ such that, for all $\gamma \in [\gamma, \overline{\gamma}]$ and $\varepsilon_C, \varepsilon_D < \varepsilon$, there are steady states with shares satisfying $|\mu^D_1 - \delta|, |\mu^P - \overline{\mu}^P|, |\mu^S - \overline{\mu}^S| \leq \eta$, and $\gamma(1 - \varepsilon_C - \varepsilon_D)(\overline{\mu}^P - \eta - (\overline{\mu}^S + \eta)g - (\delta + 2\eta)l)/(1 - \gamma + \gamma\varepsilon_C) > l$.

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

$$|\gamma^{J(\gamma, \delta)} - (1 - \delta)| \leq \frac{\eta}{6},$$

$$|\gamma^{J(\gamma, \delta)}(1 - \beta(\overline{\gamma}, 0, \delta)K(\gamma, \delta)) - \overline{\mu}^P| \leq \frac{\eta}{6},$$

$$|\gamma^{J(\gamma, \delta)}(1 - \beta(\overline{\gamma}, 0, \delta + 2(\overline{\gamma}))K(\gamma, \delta)) - \overline{\mu}^P| \leq \frac{\eta}{6},$$

$$\frac{\overline{\gamma}}{1 - \overline{\gamma}}(\overline{\mu}^P - \eta - (\overline{\mu}^S + \eta)g - (\delta + 2\eta)l) > l. \quad \text{(OA 2)}$$

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 -
Moreover, as \( \gamma \to 1 \), the left-hand side of the fourth inequality approaches infinity. The argument for the third inequality is a little more involved. Let \( K'(\gamma, \delta) = \left[ \frac{\ln(\gamma^{J(\gamma, \delta)} - \bar{p}^P)}{\ln(\beta(\gamma, 0, \delta + 2(1 - \gamma))} \right] \). It can be shown that \( \lim_{\gamma \to 1} K(\gamma, \delta)/K'(\gamma, \delta) = 1 \). Moreover, \( \lim_{\gamma \to 1} \beta(\gamma, 0, \delta + 2(1 - \gamma)) = 1 - \bar{p}^P/(1 - \delta) \), so it follows that \( \lim_{\gamma \to 1} \beta(\gamma, 0, \delta + 2(1 - \gamma))^{K'(\gamma, \delta)/K(\gamma, \delta)} = \lim_{\gamma \to 1} (\beta(\gamma, 0, \delta + 2(1 - \gamma))^{K'(\gamma, \delta)/K(\gamma, \delta)} = 1 - \bar{p}^P/(1 - \delta) \). Combining this with \( \lim_{\gamma \to 1} \gamma^{J(\gamma, \delta)} = 1 - \delta \) gives \( \lim_{\gamma \to 1} \gamma^{J(\gamma, \delta)} \left( 1 - \beta(\gamma, 0, \delta + 2(1 - \gamma))^{K'\gamma, \delta)} \right) = \bar{p}^P \).

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

\[
\begin{align*}
|\alpha(\gamma, 1 - \varepsilon_D) - (1 - \delta)| &\leq \frac{\eta}{3}, \\
|\alpha(\gamma, 1 - \varepsilon_D)(1 - \beta(\gamma, \varepsilon, \delta)\mathcal{K}) - \bar{p}^P| &\leq \frac{\eta}{3}, \\
|\alpha(\gamma, 1 - \varepsilon_D)\left( 1 - \beta(\gamma, \varepsilon, \delta + 2(1 - \gamma))\mathcal{K} \right) - \bar{p}^P| &\leq \frac{\eta}{3}, \\
\frac{\gamma(1 - \varepsilon_C - \varepsilon_D)}{1 - \gamma + \gamma\varepsilon_C} (\bar{p}^P - \eta - (\bar{p}^S + \eta)g - (\delta + 2\eta)l) &> l,
\end{align*}
\]

for all \( \gamma \in [\gamma, \bar{\gamma}] \) and \( \varepsilon_C, \varepsilon_D < \varepsilon \).

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