Online Appendix

Carrot or Stick: The Evolution of Reciprocal Preferences in a Haystack Model

by Florian Herold

Proofs

A1. *Proof of Proposition 1:*

For $0 < \gamma_+ < 1$, the following equation holds:

$$\bar{\pi}_+(\gamma_+) - \bar{\pi}_s(\gamma_+) = c_r B_{N-1,k_+^*}(\gamma_+)\left(\left(\frac{c_2 - d_2}{c_r} - 1\right) - \sum_{k=k_+^*+1}^{N-1} \frac{B_{N-1,k}(\gamma_+)}{B_{N-1,k_+^*}(\gamma_+)}\right).$$

Since, for $k > k_+^*$, $\frac{B_{N-1,k}(\gamma_+)}{B_{N-1,k_+^*}(\gamma_+)} = \left(\prod_{i=k_+^*+1}^{k} \frac{N-i}{i}\right)\left(\frac{\gamma_+}{1-\gamma_+}\right)^{(k-k_+^*)}$, it follows that $\frac{B_{N-1,k}(\gamma_+)}{B_{N-1,k_+^*}(\gamma_+)}$ is strictly increasing in γ_+, $\lim_{\gamma_+\to 0} \frac{B_{N-1,k}(\gamma_+)}{B_{N-1,k_+^*}(\gamma_+)} = 0$, and

$\lim_{\gamma_+\to 1} \frac{B_{N-1,k}(\gamma_+)}{B_{N-1,k_+^*}(\gamma_+)} = \infty$. Since $0 < \frac{c_2-d_2}{c_r} - 1 < \infty$, the difference $\bar{\pi}_+(\gamma_+) - \bar{\pi}_s(\gamma_+)$ is positive for γ_+ sufficiently close to zero and negative for γ_+ sufficiently close to one. The states $\gamma_+ = 0$ and $\gamma_+ = 1$ are therefore not stable. Furthermore, the term in the large brackets is strictly increasing in γ_+. Due to continuity, this implies a unique asymptotically stable equilibrium characterized by $\bar{\pi}_+(\gamma_+^{eq}) - \bar{\pi}_s(\gamma_+^{eq}) = 0$.

A2. *Proof of the Comparative Statics of γ_+^{eq} with N:*

We assumed $\frac{k_+^*}{N} \equiv q$ to be constant, i.e., $k_+^* = qN$ with $0 < q < 1$. We can rearrange the equilibrium Condition 4 into the following equation:

$$\text{(A1)} \quad c_2 - c_r - d_2 = c_r \sum_{k=1}^{N-1-k_+^*} \left(\left(\prod_{l=1}^{k} \frac{N-k_+^*-l}{k_+^*+l}\right)\left(\frac{\gamma_+^{eq}}{1-\gamma_+^{eq}}\right)^k\right)$$

$$\text{(A2)} \quad = c_r \sum_{k=1}^{N(1-q)-1} \left(\left(\prod_{l=1}^{k} \frac{(1-q)N-l}{qN+l}\right)\left(\frac{\gamma_+^{eq}}{1-\gamma_+^{eq}}\right)^k\right).$$

Now we prove that for constant γ_+^{eq}, the right-hand side is strictly increasing in N. Since the left-hand side is constant, γ_+^{eq} has to fall in order to equilibrate the two sides again.

The number of terms increases with N. Since all terms in equation A1 are positive, it is sufficient to prove that each term increases in N. By extending N

to real numbers, we find that:

$$\frac{\partial}{\partial N}\left(\frac{(1-q)N-l}{qN+l}\right)=\frac{l}{(qN+l)^2}>0. \tag{A3}$$

A3. Proof of Proposition 2:

For $0<\gamma_-<1$, the difference in the average fitness between a punisher and a self-interested responder can be rewritten as:

$$\bar{\pi}_-(\gamma_-)-\bar{\pi}_s(\gamma_-)=c_pB_{N-1,k_-^*}(\gamma_-)\left(\left(\frac{c_2-d_2}{c_p}\right)-\sum_{k=0}^{k_-^*-1}\frac{B_{N-1,k}(\gamma_-)}{B_{N-1,k_-^*}(\gamma_-)}\right). \tag{A4}$$

For $k<k_-^*$, it is the case that: $\frac{B_{N-1,k}(\gamma_-)}{B_{N-1,k_-^*}(\gamma_-)}=\left(\prod_{i=k+1}^{k_-^*}\frac{i}{N-i}\right)\left(\frac{1-\gamma_-}{\gamma_-}\right)^{(k_-^*-k)}$; and thus $\frac{B_{N-1,k}(\gamma_-)}{B_{N-1,k_-^*}(\gamma_-)}$ is strictly decreasing in γ_-, $\lim_{\gamma_-\to 0}\frac{B_{N-1,k}(\gamma_-)}{B_{N-1,k_-^*}(\gamma_-)}=\infty$, and $\lim_{\gamma_-\to 1}\frac{B_{N-1,k}(\gamma_-)}{B_{N-1,k_-^*}(\gamma_-)}=0$. This implies that the difference $\bar{\pi}_-(\gamma_-)-\bar{\pi}_s(\gamma_-)$ is negative for sufficiently small γ_- and positive for γ_- sufficiently close to one (note that in all terms of the sum we have $k<k_-^*$). Thus, the states $(\gamma_-=0)$ and $(\gamma_-=1)$ are both asymptotically stable. Furthermore, the term in the large brackets is strictly increasing in γ_-. Due to continuity this implies a unique interior rest point γ_-^{cut}, which is unstable and characterized by

$$\bar{\pi}_-(\gamma_-^{cut})-\bar{\pi}_s(\gamma_-^{cut})=0. \tag{A5}$$

A4. Technical Definitions and Conditions Used in Proofs of Setting 3:

Let $k_\pm^*\equiv\lfloor\frac{N(d_1-c_1)}{r+p}\rfloor$. This number of reciprocators is one too few to induce cooperation.

For $k_\pm\le k_\pm^*$, let $k_+^*(k_\pm)\equiv\lfloor\frac{N(d_1-c_1)-k_\pm(r+p)}{r}\rfloor$. For a given number of reciprocal players $k_\pm$, this number of rewarders is exactly one too few to induce proposers to cooperate.

For $k_\pm\le k_\pm^*$, let $k_-^*(k_\pm)\equiv\lfloor\frac{N(d_1-c_1)-k_\pm(r+p)}{p}\rfloor$. For a given number of reciprocal players $k_\pm$, this number of punishers is exactly one too few to induce proposers to cooperate.

We define $\mathcal{U}_{+,\pm}\equiv\{(k_+,k_\pm):k_+r+k_\pm(r+p)<N(d_1-c_1)\}$, and for any $(k_+,k_\pm)\in\mathcal{U}_{+,\pm}$, we define $k_-^*(k_+,k_\pm)\equiv\lfloor\frac{N(d_1-c_1)-k_+r-k_\pm(r+p)}{p}\rfloor$. For a given number k_+ of rewarders and a given number $k_\pm$ of reciprocal players in a group, the number $k_-^*(k_+,k_\pm)$ corresponds to the highest number of punishers which is still not sufficient to induce cooperation in that group.

Let $\mathcal{U}_{-,\pm}\equiv\{(k_-,k_\pm):k_-p+k_\pm(r+p)<N(d_1-c_1)\}$. For $(k_-,k_\pm)\in\mathcal{U}_{-,\pm}$, let $k_+^*(k_-,k_\pm)\equiv\lfloor\frac{N(d_1-c_1)-k_-p-k_\pm(r+p)}{r}\rfloor$. For a given number of punishers k_- and

reciprocal players $k_\pm$ in a group, this number of rewarders is exactly one too few to induce proposers to cooperate.

CONDITION 1: *For $i, j \in \{+, -\}$ with $i \neq j$, let the equation $k_i^*(k_\pm^*) = 0$ hold and let $k_j^*(k_i^*(k_\pm), k_\pm) = 0$ for all $k_\pm < k_\pm^*$.*

Condition 1 ensures that for any number of reciprocal players that is not yet sufficient to induce cooperation, (i) a single rewarder (jointly with a number of punishers) can potentially be pivotal in inducing cooperation and (ii) a single punisher (jointly with a number of rewarders) can be pivotal in inducing cooperation. Note that Assumption 5 implies Condition 1.

A5. Proof of Proposition 3:

Step 1: First we show that a monomorphic population of punishers is asymptotically stable. To prove this, it is sufficient to show that punishers earn strictly higher fitness than any other type when the proportion of nonpunishers in a population state is sufficiently small.[27] First consider a type who rewards after cooperation, i.e., a rewarder or a reciprocal responder. Almost all groups cooperate and he has to bear the costs c_r of rewarding with a probability arbitrarily close to one. On the other hand, he can outperform a punisher only in noncooperative groups which have a proportion arbitrarily close to zero. Hence, punishers strictly outperform rewarders if the proportion of punishers is close enough to one. Now consider the expected fitness difference between a punisher and a self-interested responder. Being a punisher increases fitness by $(c_2 - d_2)$ in groups where this has a pivotal effect, but costs c_p in groups where cooperation is not achieved. Hence, the difference in average payoffs between punishers and self-interested responders is given by the equation:

$$\begin{aligned} \bar{\pi}_-(\boldsymbol{\gamma}) \quad -\bar{\pi}_s(\boldsymbol{\gamma}) = &\sum_{(k_+, k_\pm) \in \mathcal{U}_{+,\pm}} \gamma_+^{k_+} \gamma_\pm^{k_\pm} \gamma_s^{N-1-k_+-k_\pm-k_-^*(k_+,k_\pm)} \gamma_-^{k_-^*(k_+,k_\pm)} \\ \text{(A6)} \qquad & \left[(c_2 - d_2) \binom{N-1}{k_+, k_\pm, k_-^*(k_+, k_\pm)} - c_p \sum_{k_- < k_-^*(k_+, k_\pm)} \binom{N-1}{k_+, k_\pm, k_-} \left(\frac{\gamma_s}{\gamma_-} \right)^{k_-^*(k_+,k_\pm) - k_-} \right] \end{aligned}$$

See Subsection A.A4 for the definition of $k_-^*(\cdot, \cdot)$ and related expressions. As γ_- becomes sufficiently close to one, γ_s becomes arbitrarily small and the term in the square brackets is strictly positive for all k_+. Hence, punishers strictly outperform self-interested players (and all other types) in a sufficiently small neighborhood of the monomorphic punisher population.

Step 2: The statement that both a monomorphic population of rewarders and a monomorphic population of self-interested players are unstable can be seen directly from Proposition 1. A monomorphic population of reciprocal responders is not stable since any proportion of invading punishers will be more successful since they save the costs of rewarding but still obtain full cooperation.

[27]Then Step 1 follows from, for example, Lyapunov's direct method (explained in Weibull (1995), for instance,) with Lyapunov function $L(\boldsymbol{\gamma}) \equiv (\gamma_- - 1)^2$.

A6. Proof of Proposition 4:

The following lemma is useful in proving Proposition 4.

LEMMA 1: *In Setting 3 let Condition 1 hold in addition to Assumptions 1-4. Then there exists a cost $\tilde{c}_p > 0$ such that for all $c_p \leq \tilde{c}_p$, there exist values $\underline{\gamma}_s$ and $\overline{\gamma}_s$ such that $0 < \underline{\gamma}_s < \overline{\gamma}_s < 1$, and such that the following statements hold:*

(i) $\bar{\pi}_+(\boldsymbol{\gamma}) > \bar{\pi}_s(\boldsymbol{\gamma})$ *whenever* $1 > \gamma_s \geq \underline{\gamma}_s$*;*

(ii) $\bar{\pi}_-(\boldsymbol{\gamma}) > \bar{\pi}_s(\boldsymbol{\gamma})$ *whenever* $0 < \gamma_s \leq \overline{\gamma}_s$*; and*

(iii) $\bar{\pi}_\pm(\boldsymbol{\gamma}) > \bar{\pi}_s(\boldsymbol{\gamma})$ *whenever* $\underline{\gamma}_s \leq \gamma_s \leq \overline{\gamma}_s$*.*

Proof of Lemma 1 (see Subsec. A.A4 for the definition of $k^*_-(\cdot,\cdot)$ and related expressions):

$$\begin{aligned}
&\bar{\pi}_+(\boldsymbol{\gamma}) - \bar{\pi}_s(\boldsymbol{\gamma}) \\
= \; &(c_2 - d_2 - c_r) \sum_{(k_-,k_\pm)\in\mathcal{U}_{-,\pm}} \binom{N-1}{k_-, k_\pm, k^*_+(k_-,k_\pm)} \gamma_+^{k^*_+(k_-,k_\pm)} \gamma_-^{k_-} \gamma_\pm^{k_\pm} \gamma_s^{N-1-k^*_+(k_-,k_\pm)-k_--k_\pm} \\
&-c_r \sum_{(k_-,k_\pm)\in\mathcal{U}_{-,\pm}} \sum_{k_+>k^*_+(k_-,k_\pm)} \binom{N-1}{k_-, k_\pm, k_+} \gamma_+^{k_+} \gamma_-^{k_-} \gamma_\pm^{k_\pm} \gamma_s^{N-1-k_+-k_--k_\pm} \\
&-c_r \sum_{(k_-,k_\pm)\notin\mathcal{U}_{-,\pm}} \sum_{k_+\geq 0} \binom{N-1}{k_-, k_\pm, k_+} \gamma_+^{k_+} \gamma_-^{k_-} \gamma_\pm^{k_\pm} \gamma_s^{N-1-k_+-k_--k_\pm}
\end{aligned}$$

$$\begin{aligned}
\geq \; &\sum_{(k_-,k_\pm)\in\mathcal{U}_{-,\pm}} \gamma_-^{k_-} \gamma_\pm^{k_\pm} \gamma_+^{k^*_+(k_-,k_\pm)} \gamma_s^{N-1-k_--k_\pm-k^*_+(k_-,k_\pm)} \Bigg[\frac{c_2-d_2-c_r}{2} \binom{N-1}{k_-, k_\pm, k^*_+(k_-,k_\pm)} \\
&-c_r \sum_{k_+>k^*_+(k_-,k_\pm)} \binom{N-1}{k_-, k_\pm, k_+} \left(\frac{\gamma_+}{\gamma_s}\right)^{k_+-k^*_+(k_-,k_\pm)} \Bigg] \\
+ \; &\sum_{k_\pm<k^*_\pm} \gamma_-^{k^*_-(k_\pm)} \gamma_\pm^{k_\pm} \gamma_s^{N-1-k^*_-(k_\pm)-k_\pm} \Bigg[\frac{c_2-d_2-c_r}{2} \binom{N-1}{k^*_-(k_\pm), k_\pm, 0} \\
&-c_r \sum_{k_->k^*_-(k_\pm)} \sum_{k_+\geq 0} \binom{N-1}{k_-, k_\pm, k_+} \left(\frac{\gamma_-}{\gamma_s}\right)^{k_--k^*_-(k_\pm)} \left(\frac{\gamma_+}{\gamma_s}\right)^{k_+} \Bigg] \\
+ \; &\gamma_\pm^{k^*_\pm} \gamma_s^{N-1-k^*_\pm} \Bigg[\frac{c_2-d_2-c_r}{2} \binom{N-1}{0, k^*_\pm, 0} - c_r \sum_{k_->k^*_-(k^*_\pm)} \sum_{k_+\geq 0} \binom{N-1}{k_-, k^*_\pm, k_+} \left(\frac{\gamma_-}{\gamma_s}\right)^{k_-} \left(\frac{\gamma_+}{\gamma_s}\right)^{k_+} \\
&-c_r \sum_{k_\pm>k^*_\pm} \sum_{k_-\geq 0} \sum_{k_+\geq 0} \binom{N-1}{k_-, k_\pm, k_+} \left(\frac{\gamma_-}{\gamma_s}\right)^{k_-} \left(\frac{\gamma_\pm}{\gamma_s}\right)^{k_\pm-k^*_\pm} \left(\frac{\gamma_+}{\gamma_s}\right)^{k_+} \Bigg],
\end{aligned} \tag{A7}$$

where we split the first sum of positive terms into two equal parts and then, in the second part, dropped all positive terms except those for which $k^*_+(\cdot,\cdot)$ is zero. For γ_s sufficiently close to one, the proportions γ_+, γ_-, and $\gamma_\pm$ all become arbitrarily close to zero, the negative terms after c_r go to zero, and the total terms in the square brackets are strictly positive. Hence, there is a $\underline{\gamma}_s < 1$ such that for all $\boldsymbol{\gamma} = (\gamma_s, \gamma_+, \gamma_-, \gamma_\pm)$ with $\underline{\gamma}_s < \gamma_s < 1$, it is the case that $\bar{\pi}_+(\boldsymbol{\gamma}) > \bar{\pi}_s(\boldsymbol{\gamma})$. Note that $\underline{\gamma}_s$ does not depend on the cost of punishing, c_p.

Now let $\alpha \equiv \inf_{\gamma_s = \underline{\gamma}_s}\{(\bar{\pi}_+(\boldsymbol{\gamma}) - \bar{\pi}_s(\boldsymbol{\gamma}))\}$. Since $\bar{\pi}_\pm(\boldsymbol{\gamma}) \geq \bar{\pi}_+(\boldsymbol{\gamma}) - c_p$, it follows that, for $c_p < \frac{\alpha}{2}$, the inequality $\bar{\pi}_\pm(\boldsymbol{\gamma}) > \bar{\pi}_s(\boldsymbol{\gamma}) + \frac{\alpha}{2}$ holds for all $\boldsymbol{\gamma}$ with $\gamma_s = \underline{\gamma}_s$. Since all payoffs are continuous in $\boldsymbol{\gamma}$, this extends to a neighborhood and we can select a $\overline{\gamma}_s$ such that for all $\boldsymbol{\gamma}$ with $\underline{\gamma}_s \leq \gamma_s \leq \overline{\gamma}_s$, the inequality $\bar{\pi}_\pm(\boldsymbol{\gamma}) > \bar{\pi}_s(\boldsymbol{\gamma})$ holds.

Finally, we show that, for any value $\overline{\gamma}_s$ such that $\underline{\gamma}_s < \overline{\gamma}_s < 1$, and for sufficiently small costs of punishment $c_p > 0$, the inequality $\bar{\pi}_-(\boldsymbol{\gamma}) > \bar{\pi}_s(\boldsymbol{\gamma})$ holds for all $\boldsymbol{\gamma}$ with $0 < \gamma_s < \overline{\gamma}_s$. For such $\boldsymbol{\gamma}$, at least one of the following three conditions must hold: $\gamma_- > \frac{1-\overline{\gamma}_s}{3}$ or $\gamma_+ > \frac{1-\overline{\gamma}_s}{3}$ or $\gamma_\pm > \frac{1-\overline{\gamma}_s}{3}$. For the first condition the expression after c_p in Equation A6 is bounded and, for sufficiently small cost $c_p > 0$, the result follows. Under the second condition an analogous argument applies to the following equation:

$$\begin{aligned}
&\bar{\pi}_-(\boldsymbol{\gamma}) - \bar{\pi}_s(\boldsymbol{\gamma}) \\
&= \textstyle\sum_{(k_-,k_\pm)\in\mathcal{U}_{-,\pm}} \quad \gamma_+^{k_+^*(k_-,k_\pm)} \gamma_\pm^{k_\pm} \gamma_s^{N-1-k_+^*(k_-,k_\pm)-k_\pm-k_-} \gamma_-^{k_-} \Bigg[(c_2 - d_2) \binom{N-1}{k_+^*(k_-,k_\pm), k_\pm, k_-} \\
&\qquad - c_p \sum_{k_+ < k_+^*(k_-,k_\pm)} \binom{N-1}{k_+, k_\pm, k_-} \left(\frac{\gamma_s}{\gamma_+}\right)^{k_+^*(k_-,k_\pm)-k_+} \Bigg].
\end{aligned} \tag{A8}$$

Under the third condition an analogous argument applies to this equation:

$$\begin{aligned}
&\bar{\pi}_-(\boldsymbol{\gamma}) - \bar{\pi}_s(\boldsymbol{\gamma}) \\
&= \textstyle\sum_{(k_+,k_-)\in\mathcal{U}_{+,-}} \quad \gamma_+^{k_+} \gamma_\pm^{k_\pm^*(k_+,k_-)} \gamma_s^{N-1-k_+-k_\pm^*(k_+,k_-)-k_-} \gamma_-^{k_-} \Bigg[(c_2 - d_2) \binom{N-1}{k_+, k_\pm^*(k_+,k_-), k_-} \\
&\qquad - c_p \sum_{k_\pm < k_\pm^*(k_+,k_-)} \binom{N-1}{k_+, k_\pm, k_-} \left(\frac{\gamma_s}{\gamma_\pm}\right)^{k_\pm^*(k_+,k_-)-k_\pm} \Bigg].
\end{aligned} \tag{A9}$$

Taken together, these three cases complete the proof of Lemma 1.

We now use Lemma 1 to prove Proposition 4. As a preamble, note that from any interior state, and for any regular selection dynamics, the population state does not reach the boundaries in finite time, i.e., no preference type vanishes completely in finite time.[28]

First, we show that for any initial interior state $\boldsymbol{\gamma}$ with $\gamma_s > \underline{\gamma}_s$, after a finite time we will reach a state with $\gamma_s \leq \underline{\gamma}_s$; once the inequality $\gamma_s \leq \underline{\gamma}_s$ holds, it will continue to hold forever. The latter follows from $\min\{\bar{\pi}_-(\boldsymbol{\gamma}), \bar{\pi}_+(\boldsymbol{\gamma}), \bar{\pi}_\pm(\boldsymbol{\gamma})\} > \bar{\pi}_s(\boldsymbol{\gamma})$ for all $\boldsymbol{\gamma}$ with $\underline{\gamma}_s < \gamma_s < \overline{\gamma}_s$. Consider any initial interior state $(\gamma_+^0, \gamma_-^0, \gamma_\pm^0, \gamma_s^0)$.

[28]Compare, for example, Weibull (1995) page. 141.

Then, as long as $\underline{\gamma}_s < \gamma_s$, we must have $\frac{\gamma_s^t}{\gamma_+^t} \leq \frac{\gamma_s^0}{\gamma_+^0}$, which, since $\gamma_+^t \leq 1-\gamma_s^t$, implies in particular $\gamma_s^t \leq \frac{\gamma_s^0}{\gamma_s^0+\gamma_+^0}$. Consider now the set $A \equiv \{\boldsymbol{\gamma} : \underline{\gamma}_s \leq \gamma_s \leq \frac{\gamma_s^0}{\gamma_s^0+\gamma_+^0}\}$. For all $\boldsymbol{\gamma} \in A$, it is the case that $\bar{\pi}_+(\boldsymbol{\gamma}) > \bar{\pi}_s(\boldsymbol{\gamma})$ and hence, by payoff monotonicity, the inequality $g_+(\boldsymbol{\gamma}) > g_s(\boldsymbol{\gamma})$ also holds. Since A is compact, it is the case that $\epsilon \equiv \inf_{\boldsymbol{\gamma}\in A}\{g_+(\boldsymbol{\gamma}) - g_s(\boldsymbol{\gamma})\} > 0$. Hence, for all $t \geq 0$ such that $\boldsymbol{\gamma}^t \in A$, it follows that:

$$\text{(A10)} \quad \frac{d}{dt}\left(\frac{\gamma_s^t}{\gamma_+^t}\right) = \frac{\dot{\gamma}_s^t}{\gamma_+^t} - \frac{\gamma_s^t\dot{\gamma}_+^t}{\left(\gamma_+^t\right)^2} = \left(g_s\left(\boldsymbol{\gamma}^t\right) - g_+\left(\boldsymbol{\gamma}^t\right)\right)\frac{\gamma_s^t}{\gamma_+^t} \leq -\epsilon\frac{\gamma_s^t}{\gamma_+^t}.$$

From equation (A10) it follows that $\frac{\gamma_s^t}{\gamma_+^t} \leq \frac{\gamma_s^0}{\gamma_+^0}e^{-\epsilon t}$ and therefore $\gamma_s^t \leq \frac{\gamma_s^0}{\gamma_+^0}e^{-\epsilon t}$ for all $t \geq 0$ such that $\gamma^t \in A$.[29] It then follows directly that for all $t \geq -\frac{1}{\epsilon}\ln\left(\frac{\underline{\gamma}_s}{\gamma_s^0}\gamma_+^0\right)$, it must be the case that $\gamma_s^t \leq \underline{\gamma}_s$.

Second, for any arbitrarily small number ε with $\underline{\gamma}_s > \varepsilon > 0$, we can find a time $\tilde{t}$ such that $\gamma_s^t < \varepsilon$ holds for all times $t > \tilde{t}$. This follows from an argument completely analogous to the previous one, when we consider the compact set $B \equiv \{\boldsymbol{\gamma} : \varepsilon \leq \gamma_s \leq \underline{\gamma}_s\}$ and note that $\inf_{\boldsymbol{\gamma}\in B}\{g_-(\boldsymbol{\gamma}) - g_s(\boldsymbol{\gamma})\} > 0$.

Third, we complete the proof of Proposition 4 by showing that γ_+^t and $\gamma_\pm^t$ also converge to zero, and hence γ_- converges to one. It suffices to show that there is an $\varepsilon > 0$ such that for all $\boldsymbol{\gamma}$ with $\gamma_s \leq \varepsilon$ (again a compact set), the inequality $\overline{\pi}_-(\boldsymbol{\gamma}) > \max\{\overline{\pi}_+(\boldsymbol{\gamma}), \overline{\pi}_\pm(\boldsymbol{\gamma})\}$ holds. We find a lower bound for $\overline{\pi}_-(\boldsymbol{\gamma}) - \max\{\overline{\pi}_+(\boldsymbol{\gamma}), \overline{\pi}_\pm(\boldsymbol{\gamma})\}$ by distinguishing only between groups in which all $N-1$ other players (excluding the one under consideration) are either rewarders or punishers and groups in which at least one player is self-interested. In the first type of group, punishers earn $c_r > 0$ more than rewarders, while in the other groups the fitness advantage of rewarders is bounded (e.g., by $(c_2 - d_2 + c_p)$). When γ_s goes to zero, the total probability mass on the first type of group goes to one and the total probability mass on the second type of group goes to zero. Hence, $\overline{\pi}_-(\boldsymbol{\gamma}) > \max\{\overline{\pi}_+(\boldsymbol{\gamma}), \overline{\pi}_\pm(\boldsymbol{\gamma})\}$ for sufficiently small γ_-; this completes the third step and the proof of Proposition 4.

Additional Material

B1. Preferences in Player-Position 1

We briefly generalize the main model in a way that allows different preferences conditional on the player's position. This allows us to formally state and prove Lemma 2, which establishes that we can expect self-interested behavior in player position 1.

[29] See Weibull (1995), p. 146 for a related formal argument in the context of showing that the share of a strictly dominated strategy converges to zero for payoff-monotonic dynamics.

Let there be a single infinite population of players.[30] Let $\Theta_i \subset [0,1]^{|O|}$ (modulus affine transformations) denote the set of all possible preference types a player can have in player position $i \in \{1,2\}$. For $\theta_i \in \Theta_i$, $\theta_i(o)$ represents the subjective utility that a player in position i of type θ_i assigns to an outcome $o \in O$. Let $\Theta \equiv \Theta_1 \times \Theta_2$ be the set of all preference types. We restrict our analysis to finite sets of preference types and we assume that each preference type in each role has strict preferences over outcomes.

For a specific group let k_{θ_i} be the number of players of type $\theta_i \in \Theta_i$ in player position i. Thus, for any $i \in \{1,2\}$, the equation $\sum_{\theta_i \in \Theta_i} k_{\theta_i} = N$ holds. Let $\mathbf{k_i}$ be the vector of the k_{θ_i} and thus $\mathbf{k} \equiv (\mathbf{k_1}, \mathbf{k_2})$ is a vector representing the entire preference type composition of a group.

ASSUMPTION 6: **Distribution of Groups in the Haystack Model** *Each group member in each player position $i \in \{1,2\}$ is drawn independently from the infinite player population with the probability of drawing type $\theta_i \in \Theta_i$ equal to the proportion γ_{θ_i} of that type in the population. The probability that a group has composition $\mathbf{k}$ if the population state is $\boldsymbol{\gamma}$ is thus given by*

$$B_{\mathbf{k}}(\boldsymbol{\gamma}) = \prod_{i \in \{1,2\}} \left(\frac{N!}{\prod_{\theta_i \in \Theta_i} k_{\theta_i}!} \prod_{\theta_i \in \Theta_i} \left(\gamma_{\theta_i}\right)^{k_{\theta_i}} \right). \tag{B1}$$

The following shows that we should expect fitness-maximizing behavior of players in position 1.

LEMMA 2: *Let $\Theta \equiv \Theta_1 \times \Theta_2$ where Θ_1 contains the purely self-interested, fitness-maximizing preference type which has (in the role of the proposer) for each outcome a subjective utility that corresponds to the proposer's fitness payoff. Under Assumption 6 and Assumption 4 , it follows that:*

(a) The expected fitness of any preference type is weakly dominated by the corresponding preference type which has identical preferences in player-position 2, but self-interested preferences in player-position 1.

(b) A preference type whose behavior as proposer is inconsistent with expected fitness maximization in any group obtains in every interior population state $\boldsymbol{\gamma} \in int(\Gamma)$ a strictly lower expected fitness level than the corresponding preference type with identical preferences in player-position 2, but self-interested preferences in player-position 1.

(c) Any preference type (θ_1, θ_2) in the support of any stationary state $\boldsymbol{\gamma} \in \Gamma$ that also contains a positive proportion of the corresponding type (θ_s, θ_2) must act as a proposer consistent with fitness maximization in any group that occurs with positive probability.

[30] All results can be derived equally in a model of two infinite populations, one proposer population and one responder population.

(d) *In any asymptotically stable state $\boldsymbol{\gamma} \in \Gamma$, all preference types in the support of $\boldsymbol{\gamma}$ must act as a proposer consistent with fitness maximization in all possible groups.*

We proof Lemma 2 only in the context of our haystack model. Proving the corresponding statements in the model of a single infinite population without group structure is completely analogous and thus omitted.

B2. Proof of Lemma 2:

Lemma 2 follows from two observations. First, by Assumption 6, the distribution of responders' types that a proposer faces in his group is independent of his own type. Second, responders condition their responses only on the proposer's action but not on his type or the distribution of types in their group. This holds because players have preferences over outcomes only, and a responder's choosing an action is equivalent to his choosing an outcome of the game. In other words, the response profile which a proposer faces conditional on his own action is independent of his own type. Note also that the preference type a player has in position 1 has no influence on his fitness when he acts in position 2 because (i) the type distribution of the other players in his group is independent of his type and because (ii) proposers have preferences over outcomes only. Since a proposer's self-interested preferences induce fitness-maximizing behavior for any given response profile, a preference type $\theta = (\theta_s, \theta_2)$ with self-interested preferences when acting in position 1 must weakly dominate any preference type $\theta' = (\theta_{1s}, \theta_2)$ which has the identical preferences θ_2 when in position 2, and differs only in his type only when in position 1; this proves Part (a). Part (b) also follows immediately, given the fact that under Assumption 6, for every interior state, every group composition has a strictly positive probability.

Part (c) follows since a a player of type (θ_1, θ_2) who deviates in position 1 from fitness maximization in some groups that occur with positive probability will earn strictly less fitness than a player of type (θ_{1s}, θ_2) would earn if he replaced him in these groups; and he will earn weakly less in all other groups. Thus, type (θ_{1s}, θ_2) earns a strictly higher expected fitness and must have a higher growth rate by payoff monotonicity; this contradicts stationarity.

We prove Part (d) by contradiction. Suppose there is an asymptotically stable state $\boldsymbol{\gamma}$ with a strictly positive proportion γ_θ of a type $\theta = (\theta_1, \theta_2)$, where proposers of type θ_1 deviate from fitness-maximizing behavior for some group composition. Let $\theta' \equiv (\theta_s, \theta_2)$ denote the corresponding type who has self-interested preferences when in the role of the proposer, and identical preferences θ_2 when in the role of the responder. Consider any neighborhood of $\boldsymbol{\gamma}$. It must contain a state $\tilde{\boldsymbol{\gamma}}$ which is an interior state with $\tilde{\gamma}_\theta < \gamma_\theta$ and $\tilde{\gamma}_{\theta'} > \gamma_{\theta'}$. It is well-known that the solution trajectory starting from any interior state will always stay in the interior of Γ; by Part (b) and payoff monotonicity, this implies that for any time $\tau > 0$, the inequalities $\frac{\xi_{\theta'}(\tau,\tilde{\gamma})}{\xi_\theta(\tau,\tilde{\gamma})} > \frac{\tilde{\gamma}_{\theta'}}{\tilde{\gamma}_\theta} > \frac{\gamma_{\theta'}}{\gamma_\theta}$ hold; this contradicts convergence to state $\boldsymbol{\gamma}$.

B3. Details and Formal Statement of Robustness Checks

In this appendix we formally define the concept of a Δ-bounded perturbation of the evolutionary process, which is general enough to encompass all three robustness checks. In such a perturbed process players do not always play optimally. This encompasses settings where players tremble occasionally, or where players have to learn to play optimally in their respective groups. We also allow that a small number of groups may change their type composition in subperiods between the reshuffling of groups and the probability distribution changes correspondingly. This is useful for the analysis of within-group evolutionary effects.

The results we derived for our haystack model turn out to be robust to perturbations which are bounded by a sufficiently small $\Delta > 0$ in the following sense. A small interval containing any state that was asymptotically stable in the unperturbed model is an asymptotically stable set in the perturbed model. Furthermore, while new asymptotically stable states or sets can potentially arise in the perturbed model, the basins of attraction of these newly emerged stable states vanish as Δ becomes arbitrarily small. The basin of attraction of the asymptotically stable sets around the stable equilibria from the unperturbed model do not vanish with small Δ. In summary, in all three settings our previous results change only slightly under such Δ-bounded perturbations, provided the perturbations are sufficiently small.

Let $P^t_{\theta,\mathbf{k}}(\boldsymbol{\gamma})$ denote the probability in subperiod t that a player of type θ is in a group that has the constant composition $\mathbf{k}$ in all T subperiods. Similarly, let $B_{\theta,\mathbf{k}}(\boldsymbol{\gamma})$ denote the probability that a player of type θ ends up in a group of composition $\mathbf{k}$ where the distribution over groups is Multinomial as in Assumption 2.[31]

Fix a setting and the corresponding stage game G. For each group with N players in either positions and of composition $\mathbf{k}$ let $a^t_{\mathbf{k}}$ denote the vector of actions that each member in this group takes at time t given the perturbed process. Let $a^*_{\mathbf{k}}$ denote the vector of actions played in each subperiod in the main model in Section I (including Assumption 2). Without loss of generality we assume that all such vectors are sorted in the same sequence of types. Let $W^t_{\mathbf{k}}$ be the probability distribution over action-profiles in subperiod t which is generated by the perturbed process given a group which has the constant type composition $\mathbf{k}$ in all T subperiods.

DEFINITION 1 (Δ-bounded perturbation of the evolutionary process):

Let $\Delta > 0$. For each of the considered settings we call an evolutionary process a Δ-bounded perturbation of the main model in Section I if for all subperiods

[31] This probability corresponds to the probability from the Multinomial distribution for groups of size $N-1$ for a group of composition that differs from $\mathbf{k}$ only by reducing the number of players of type θ by one.

$t \in \{1, \ldots, T\}$, all types θ, all population states $\boldsymbol{\gamma} \in \Gamma$ and each possible group composition $\mathbf{k}$ it holds that $\left|P^t_{\theta,\mathbf{k}}(\boldsymbol{\gamma}) - B_{\theta,\mathbf{k}}(\boldsymbol{\gamma})\right| < \Delta$ and if, conditional on being in a group which has composition $\mathbf{k}$ in all T periods, for all $t > \Delta T$ it holds that $W^t_{\mathbf{k}}(a^t_{\mathbf{k}} = a^*_{\mathbf{k}}) \geq (1 - \Delta)$.

DEFINITION 2 (δ-stable set): *Let $\delta > 0$. We call an asymptotically stable set $\mathcal{S}$* **δ-stable** *if all $\boldsymbol{\gamma}$ with $d(\boldsymbol{\gamma}, \mathcal{S}) < \delta$ are in the basin of attraction of $\mathcal{S}$.*

Note that on the one hand for sufficiently small $\delta > 0$ all stable equilibria derived in the previous sections are δ-stable.[32] On the other hand, for any given $\delta > 0$ the newly arising stable states of the perturbed model will not be δ-stable if the perturbation becomes sufficiently small.

PROPOSITION 5: *For each setting there exists a $\overline{\delta} > 0$ such that for all $\delta \in (0, \overline{\delta})$ there is a $\Delta > 0$ such that for all Δ-bounded perturbations of the main model holds:*

in Setting 1 *$[\gamma^{eq}_+ - \delta, \gamma^{eq}_+ + \delta]$ is a δ-stable set and there exists no δ-stable set which is disjoint from $[\gamma^{eq}_+ - \delta, \gamma^{eq}_+ + \delta]$.*

in Setting 2 *the set $[0, \delta]$ and the set $[1 - \delta, 1]$ are δ-stable. There exists no δ-stable set which is disjoint from both of these sets.*

in Setting 3 *the set $\{\boldsymbol{\gamma} : \gamma_- \geq 1 - \delta\}$ is δ-stable. Furthermore, for $\theta \in \{+, \pm, s\}$ the sets $\{\boldsymbol{\gamma} : \gamma_\theta \geq 1 - \delta\}$ are not δ-stable. If, in addition, Condition 1 holds, and if the cost of punishing $c_p > 0$ are sufficiently small, then there exists no δ-stable set which is disjoint from $\{\boldsymbol{\gamma} : \gamma_- \geq 1 - \delta\}$. The population converges to this δ-stable set from any interior state γ with $\gamma_s < 1 - \delta$.*

B4. Proof of Proposition 5:

Step 1: Fix a setting. Let $\overline{\pi}_{i,\Delta}(\boldsymbol{\gamma})$ denote the expected average fitness of type i in state $\boldsymbol{\gamma} \in \Gamma$ in the Δ-bounded perturbation of the model. Then, for every $\epsilon > 0$ there exists a $\Delta > 0$ such that for all $\boldsymbol{\gamma} \in \Gamma$ and all $\theta, \theta' \in \{+, -, s\}$ holds

$$\left|\left(\overline{\pi}_\theta(\boldsymbol{\gamma}) - \overline{\pi}_{\theta'}(\boldsymbol{\gamma})\right) - \left(\overline{\pi}_{\theta,\Delta}(\boldsymbol{\gamma}) - \overline{\pi}_{\theta',\Delta}(\boldsymbol{\gamma})\right)\right| < \frac{\epsilon}{2}. \tag{B2}$$

To show this, let $b \equiv c_2 - d_2 + c_p$ denote the maximal possible per subperiod difference in material payoffs for responders. Let $\mathcal{K}$ denote the set of all possible group compositions $\mathbf{k}$ and thus $|\mathcal{K}|$ denotes the number of different $\mathbf{k} \in \mathcal{K}$. We show that for all $\theta \in \{+, -, s\}$, $|\overline{\pi}_\theta(\boldsymbol{\gamma}) - \overline{\pi}_{\theta,\Delta}(\boldsymbol{\gamma})| < (2|\mathcal{K}| + 1)b\Delta$, and thus equation (B2) follows with $\Delta \equiv \frac{\epsilon}{4(2|\mathcal{K}|+1)b}$ from the triangular inequality.

Let $\tilde{P}^t_{\theta,\mathbf{k}}(\boldsymbol{\gamma})$ denote the probability that a player of type θ is in a group that has the constant composition $\mathbf{k}$ in all T subperiods and in subperiod t the action

[32] E.g. in Setting 1 $\delta < \min\{\gamma^{eq}_+, 1 - \gamma^{eq}_+\}$ works, in Setting 2 $\delta < \min\{\gamma^{cut}_-, 1 - \gamma^{cut}_-\}$.

profile played in this group is $a^*_{\mathbf{k}}$. Thus, $\tilde{P}^t_{\theta,\mathbf{k}}(\boldsymbol{\gamma}) = P_{\theta,\mathbf{k}}(\boldsymbol{\gamma})W^t_{\mathbf{k}}(a^t_{\mathbf{k}} = a^*_{\mathbf{k}})$. For all $t > \Delta T$ we have

$$\begin{aligned}\left|\tilde{P}^t_{\theta,\mathbf{k}}(\boldsymbol{\gamma}) - B_{\theta,\mathbf{k}}(\boldsymbol{\gamma})\right| &= \left|P_{\theta,\mathbf{k}}(\boldsymbol{\gamma})W^t_{\mathbf{k}}(a^t_{\mathbf{k}} = a^*_{\mathbf{k}}) - B_{\theta,\mathbf{k}}(\boldsymbol{\gamma})\right| \\ &= \left|\left(P_{\theta,\mathbf{k}}(\boldsymbol{\gamma}) - B_{\theta,\mathbf{k}}(\boldsymbol{\gamma})\right)W^t_{\mathbf{k}}(a^t_{\mathbf{k}} = a^*_{\mathbf{k}}) - B_{\theta,\mathbf{k}}(\boldsymbol{\gamma})\left(1 - W^t_{\mathbf{k}}(a^t_{\mathbf{k}} = a^*_{\mathbf{k}})\right)\right| \\ &\leq \left|P_{\theta,\mathbf{k}}(\boldsymbol{\gamma}) - B_{\theta,\mathbf{k}}(\boldsymbol{\gamma})\right| + B_{\theta,\mathbf{k}}(\boldsymbol{\gamma})\left|1 - W^t_{\mathbf{k}}(a^t_{\mathbf{k}} = a^*_{\mathbf{k}})\right| < 2\Delta. \qquad \text{(B3)}\end{aligned}$$

and thus

$$\begin{aligned}\left|\overline{\pi}_\theta(\boldsymbol{\gamma}) - \overline{\pi}_{\theta,\Delta}(\boldsymbol{\gamma})\right| &\leq b\Delta + (1-\Delta)\left(\sum_{\mathbf{k}\in\mathcal{K}}\left|\tilde{P}^t_{\theta,\mathbf{k}}(\boldsymbol{\gamma}) - B_{\theta,\mathbf{k}}(\boldsymbol{\gamma})\right| b\right) \\ &< b\Delta + (1-\Delta)\, b2|\mathcal{K}|\Delta \leq (2|\mathcal{K}|+1)\, b\Delta. \qquad \text{(B4)}\end{aligned}$$

Step 2: Setting 1: Let $0 < \overline{\delta} < \frac{1}{4}\min\{\gamma^{eq}_+, \left(1-\gamma^{eq}_+\right)\}$. Consider any $\delta \in (0, \overline{\delta})$. Let $\epsilon \equiv \min_{\gamma_+\in[\frac{\delta}{2},\gamma^{eq}_+-\frac{\delta}{2}]\cup[\gamma^{eq}_++\frac{\delta}{2},1]}\left(\left|\overline{\pi}_+(\boldsymbol{\gamma}) - \overline{\pi}_s(\boldsymbol{\gamma})\right|\right) > 0$. If we choose $\Delta > 0$ sufficiently small, then, by step 1, $\left(\overline{\pi}_{+,\Delta}(\boldsymbol{\gamma}) - \overline{\pi}_{s,\Delta}(\boldsymbol{\gamma})\right) > 0$ on $[\frac{\delta}{2}, \gamma^{eq}_+ - \frac{\delta}{2}]$ and $\left(\overline{\pi}_{+,\Delta}(\gamma) - \overline{\pi}_{s,\Delta}(\gamma)\right) < 0$ on $[\gamma^{eq}_+ + \frac{\delta}{2}, 1]$. Hence, $[\gamma^{eq}_+ - \delta, \gamma^{eq}_+ - \delta]$ is a δ-stable set and no disjoint set can be δ-stable. Note that $\{\boldsymbol{\gamma} : \gamma_+ = 0\}$ is asymptotically stable, but not δ-stable since the basin of attraction is smaller than $\frac{\delta}{2}$.

Setting 2: Let $0 < \overline{\delta} < \frac{1}{4}\min\{\gamma^{cut}_-, \left(1-\gamma^{cut}_-\right)\}$. Consider any $\delta \in (0, \overline{\delta})$. Let $\epsilon \equiv \min_{\gamma_-\in[\frac{\delta}{2},\gamma^{cut}_--\frac{\delta}{2}]\cup[\gamma^{cut}+\frac{\delta}{2},1]}\left(\left|\overline{\pi}_-(\boldsymbol{\gamma}) - \overline{\pi}_s(\boldsymbol{\gamma})\right|\right) > 0$. If we choose $\Delta > 0$ sufficiently small, according to step 1, then $\left(\overline{\pi}_{-,\Delta}(\gamma) - \overline{\pi}_{s,\Delta}(\gamma)\right) < 0.$ on $[0, \gamma^{cut}_- - \frac{\delta}{2}]$ and $\left(\overline{\pi}_{-,\Delta}(\boldsymbol{\gamma}) - \overline{\pi}_{s,\Delta}(\boldsymbol{\gamma})\right) > 0$ on $[\gamma^{cut}_- + \frac{\delta}{2}, 1 - \frac{\delta}{2}]$. Hence, $[0, \delta]$ is a δ-stable set and so is $[1-\delta, 1]$. There cannot exist any further δ-stable set. Note in particular, that any small set around the equilibrium state(s) in $[\gamma^{cut}_- - \frac{\delta}{2}, \gamma^{cut}_- + \frac{\delta}{2}]$ is not δ stable since, for instance at the upper value of the boundary plus δ, γ_- must be increasing.

Setting 3: (the steps are analogous to the proof of Proposition 3 and 4.)

Step i): From step 1 of the proof of Prop. 3 we know that there is a $\hat{\delta} > 0$ such that for all $\boldsymbol{\gamma}$ with $1 > \gamma_- \geq 1 - \hat{\delta}$ holds $\overline{\pi}_-(\boldsymbol{\gamma}) - \max\{\overline{\pi}_s(\boldsymbol{\gamma}), \overline{\pi}_+(\boldsymbol{\gamma})\} > 0$. Let $\delta < \frac{\hat{\delta}}{4}$. Thus, $\epsilon \equiv \min_{\boldsymbol{\gamma}\in\{\boldsymbol{\gamma}:1-3\delta\leq\gamma_-\leq1-\delta\}}\left(\overline{\pi}_-(\boldsymbol{\gamma}) - \max\{\overline{\pi}_s(\boldsymbol{\gamma}), \overline{\pi}_+(\boldsymbol{\gamma})\}\right) > 0$ and hence, again by step 1, we can find a $\Delta > 0$ such that the inequality $\left(\overline{\pi}_{-,\Delta}(\boldsymbol{\gamma}) - \max\{\overline{\pi}_{s,\Delta}(\boldsymbol{\gamma}), \overline{\pi}_{+,\Delta}(\boldsymbol{\gamma})\}\right) > 0$ holds for all $\boldsymbol{\gamma}$ in $\{\boldsymbol{\gamma} : 1 - 3\delta \leq \gamma_- \leq 1 - \delta\}$, and thus $\{\boldsymbol{\gamma} : \gamma_- \geq 1 - \delta\}$ is δ-stable.

Step ii): From Setting 1 and Setting 2 it follows immediately that for $\theta \in \{+, s\}$ the sets $\{\boldsymbol{\gamma} : \gamma_\theta \geq 1 - \delta\}$ are not δ-stable.

Step iii): For $\delta > 0$ and sufficiently small $c_p > 0$ we can find $\underline{\gamma}_s$ and $\overline{\gamma}_s$ with $0 < \underline{\gamma}_s < 1 - \delta < \overline{\gamma}_s < 1$ such that $\bar{\pi}_+(\boldsymbol{\gamma}) > \bar{\pi}_s(\boldsymbol{\gamma})$ whenever $1 > \gamma_s \geq \underline{\gamma}_s$, $\bar{\pi}_-(\boldsymbol{\gamma}) > \bar{\pi}_s(\boldsymbol{\gamma})$ whenever $0 < \gamma_s \leq \overline{\gamma}_s$, and $\bar{\pi}_\pm(\boldsymbol{\gamma}) > \bar{\pi}_s(\boldsymbol{\gamma})$ whenever $\underline{\gamma}_s \leq \gamma_s \leq \overline{\gamma}_s$. Furthermore we know from the last paragraph of step 3 of the proof of Proposition 4 that there is an $\tilde{\gamma}_s > 0$ such that for all $\boldsymbol{\gamma}$ with $\gamma_s < \tilde{\gamma}_s$ holds $\overline{\pi}_-(\boldsymbol{\gamma}) >$

$\max\{\overline{\pi}_+(\boldsymbol{\gamma}), \overline{\pi}_\pm(\boldsymbol{\gamma})\} > 0$. Let $\epsilon_1 \equiv \min_{\gamma_s \in [\underline{\gamma}_s, 1-\delta]} \{(\max\{\bar{\pi}_+(\boldsymbol{\gamma}), \bar{\pi}_\pm(\boldsymbol{\gamma})\} - \bar{\pi}_s(\boldsymbol{\gamma}))\}$, $\epsilon_2 \equiv \min_{\gamma_s \in [\frac{\min\{\tilde{\gamma}_s, \underline{\gamma}_s, \delta\}}{2}, \overline{\gamma}_s]} \{(\bar{\pi}_-(\boldsymbol{\gamma}) - \bar{\pi}_s(\boldsymbol{\gamma}))\}$, $\epsilon_3 \equiv \min_{\gamma_s \in [0, \tilde{\gamma}_s]} \{(\bar{\pi}_-(\boldsymbol{\gamma}) - \max\{\bar{\pi}_+(\boldsymbol{\gamma}), \bar{\pi}_\pm(\boldsymbol{\gamma})\})\}$, and $\epsilon \equiv \min\{\epsilon_1, \epsilon_2, \epsilon_3\}$. Step 1 guaranties that for this $\epsilon > 0$ we can choose a $\Delta > 0$ such that the payoff structure that we need for step 3 of the proof of Proposition 4 is preserved for Δ-bounded perturbations and the argument there still applies.

Small perturbations of play, formal part

We allow the probability $\varepsilon_i(\mathbf{k})$ of the mistakes to depend on the player's position $i \in \{1, 2\}$, his type θ, the number of periods t he played in the group already, and the composition $\mathbf{k}$ of the respective group he is in. We only require that all this mistakes are bounded by ϵ, i.e. $\varepsilon^t_{i\theta}(\mathbf{k}) < \epsilon$ for all $i \in \{1, 2\}$, $\theta \in \Theta$, $t \in \{1, \ldots, T\}$, and $\mathbf{k}$.

One issue that arises is whether small probability trembles of players in one position do change the optimal response of the players in the other position - then behavior in a group might change significantly even in instances where no tremble occurs. Trembles by proposers cannot have this effect since responders do observe the proposers action, know at which decision node they are, and choose their optimal action correspondingly. In particular, in instances where the proposer did not tremble, responders' optimal choices match exactly their behavior from the corresponding setting without trembles. The effect of trembles by responders could however have the effect that optimal behavior of the proposers changes in groups that are close to the threshold. Yet it turns out that for sufficiently small trembles this cannot happen and the optimal response for proposers remains unaffected by the trembling of the responders.[33]

LEMMA 3: *Let $\Delta > 0$. Fix the setting. Then there is an $\epsilon > 0$ such that if $\varepsilon^t_{i\theta}(\mathbf{k}) < \epsilon$ for all $i \in \{1, 2\}$, $\theta \in \Theta$, $t \in \{1, \ldots, T\}$, and $\mathbf{k}$, then the model with such small trembles is a Δ-bounded perturbation of the main model in Section I.*

B5. Proof of Lemma 3:

$|P_{\theta,\mathbf{k}}(\gamma) = B_{\theta,\mathbf{k}}(\gamma)| < \Delta$ holds automatically since the composition of types in a group is not changed in this robustness check and therefore $P_{\theta,\mathbf{k}}(\gamma) = B_{\theta,\mathbf{k}}(\gamma)$. Furthermore, we assumed the generic case where parameters in each of the settings without mistakes are such that the expected payoff for Player 1 after one action is always strictly different from the other actions. Let g denote the minimum of the absolute value of this payoff difference for all group compositions. Payoffs are bounded. Let b denote the maximal payoff difference. If the mistake probabilities of responders are bounded by $\tilde{\epsilon}$, then, for a given action, the change

[33] Furthermore, while for intermediate trembling probabilities behavior in groups close to the threshold may change, our analysis essentially goes still through with the modification that the threshold k^* needs to be adapted accordingly.

in expected payoff for proposers is bounded by $\tilde{\epsilon}b$ and if we choose $\tilde{\epsilon} < \frac{g}{2b}$ the proposers' optimal choice of action is not influenced by these small enough mistake probabilities. Then we have for $0 < \epsilon < \frac{g}{2b}$ that $W^t_{\mathbf{k}}(a^t_{\mathbf{k}} = a^*_{\mathbf{k}}) \geq (1-\epsilon)^{2N}$ and thus for $\epsilon < \min\{\frac{g}{2b}, 1-(1-\Delta)^{\frac{1}{2N}}\}$ the inequality $W^t_{\mathbf{k}}(a^t_{\mathbf{k}} = a^*_{\mathbf{k}}) > (1-\Delta)$ holds, and thus the process is a Δ-bounded perturbation.

Example: Proposers tremble with probability ϵ_1, responders with probability ϵ_2

Setting 1: Suppose proposers play their optimal actions only with probability $(1-\epsilon_1)$ and with a small probability $\epsilon_1 > 0$ they play the other action. Then the expected payoff of a self-interested responder is

$$\sum_{k=0}^{k^*_+} \left((1-\epsilon_1)\, d_2 + \epsilon_1 c_2\right) B_{N-1,k}\left(\gamma_+\right) + \sum_{k=k^*_+ +1}^{N-1} \left(\epsilon_1 d_2 + (1-\epsilon_1)\, c_2\right) B_{N-1,k}\left(\gamma_+\right),$$

and

$$\sum_{k=0}^{k^*_+ -1} \left((1-\epsilon_1)\, d_2 + \epsilon_1 \left(c_2 - c_r\right)\right) B_{N-1,k}\left(\gamma_+\right) + \sum_{k=k^*_+}^{N-1} \left(\epsilon_1 d_2 + (1-\epsilon_1)\left(c_2 - c_r\right)\right) B_{N-1,k}\left(\gamma_+\right)$$

is the expected payoff of a rewarder. The expected difference in fitness between rewarders and self-interested responders is

$$(1-2\epsilon_1)\left(\left(c_2 - c_r - d_2\right) B_{N-1,\tilde{k}^*_+}\left(\gamma_+\right) - c_r \sum_{k_+ = \tilde{k}^*_+ +1}^{N-1} B_{N-1,k_+}\left(\gamma_+\right)\right) - \epsilon_1 c_r,$$

which corresponds the fitness difference without trembles, compressed by the factor $(1-2\epsilon_1)$, and then shifted by $-\epsilon_1 c_r$.

If, in addition, also responders tremble, play their intended action only with probability $(1-\epsilon_2)$, and the remaining action with small probability $\epsilon_2 > 0$, the calculation requires slight further adjustments. First, for larger ϵ_2 the proposers' optimal behavior may change in groups close to the threshold k^*_+ and we might have a slightly different threshold $\tilde{k}^*_+$. Under the assumption that $N\frac{d_1 - c_1}{r}$ is not an integer however, we have $\tilde{k}^*_+ = k^*_+$ for sufficiently small $\epsilon_2 > 0$. The expected difference in fitness between rewarders and self-interested responders in Setting 1 is then given by $\bar{\pi}_+(\gamma_+) - \bar{\pi}_s(\gamma_+) =$

$$(1-2\epsilon_2)\left((1-2\epsilon_1)\left(\left(c_2 - c_r - d_2\right) B_{N-1,\tilde{k}^*_+}\left(\gamma_+\right) - c_r \sum_{k_+ = \tilde{k}^*_+ +1}^{N-1} B_{N-1,k_+}\left(\gamma_+\right)\right) - \epsilon_1 c_r\right), \tag{B5}$$

i.e. by a further compression of the previous fitness difference by a factor

$(1 - 2\epsilon_2)$, which will not change which type is getting the higher expected fitness.

Setting 2: Similarly, the expected fitness difference between punishers and self-interested responders in Setting 2 becomes $\bar{\pi}_-(\gamma_-) - \bar{\pi}_s(\gamma_-) =$

(B6)

$$(1-2\epsilon_2)\left((1-2\epsilon_1)\left((c_2-d_2)\,B_{N-1,\tilde{k}_-^*}\left(\gamma_-\right) - c_p \sum_{k_-=0}^{\tilde{k}_-^*-1} B_{N-1,k_-}\left(\gamma_-\right)\right) - \epsilon_1 c_p\right).$$

Learning the distribution of preferences in a group, formal part

Here we show formally that for small $\epsilon > 0$ the learning by valuation model is a Δ-bounded perturbation of the main model. Δ becomes small when the number of subperiods T in which players interact before reshuffling gets large.

After a group is randomly formed, players' preferences remain constant until the group is reshuffled after a period of time $\mathfrak{T}$. Given their subjective preferences, players' behavior adjusts fast to optimal behavior and an equilibrium emerges. This equilibrium is played for a while before the group dissolves. We divide this time period $\mathfrak{T}$ into T subperiods of length $\frac{\mathfrak{T}}{T}$. In each subperiod the basic game is played once. We are interested in the limit when T becomes large. Then there is enough time for players to learn to play an equilibrium in the early periods and to play this equilibrium afterwards for most of the time period $\mathfrak{T}$.

LEMMA 4: *Fix a setting. For every $\Delta > 0$ and for sufficiently small $\epsilon > 0$ there exists a $T_{\epsilon,\Delta} \in \mathbb{N}$ such that for all $T > T_{\epsilon,\Delta}$ the ϵ-reinforcement learning process (which replaces Assumption 3 of an observable group composition) is a Δ-bounded perturbation of the main model in Section I.*

B6. Proof of Lemma 4:

$|P_{\theta,\mathbf{k}}(\boldsymbol{\gamma}) = B_{\theta,\mathbf{k}}(\boldsymbol{\gamma})| < \Delta$ holds automatically since the composition of types in a group is not changed in this robustness check, thus $P_{\theta,\mathbf{k}}(\boldsymbol{\gamma}) = B_{\theta,\mathbf{k}}(\boldsymbol{\gamma})$.

At a time t we call the action that scored the highest average subjective utility up to that time the "intended" action of a player. The other action with a lower score is called the "unintended action". Let $V_t(C)$ denote the proposer's valuation of action C at time t, i.e. the average payoff Player 1 received until time t in periods where he player C. Correspondingly, $V_t(D)$ denotes his valuation of the action D. Let $\Delta > 0$. Define $\tilde{\Delta} \equiv min_{k_+,k_-}\left|c_1 - d_1 + \frac{k_+}{N_2}r + \frac{k_-}{N_2}p\right| > 0$, with $k_- = 0$ in Setting 1 and with $k_+ = 0$ in Setting 2. Consider an ϵ-reinforcement learning process with $0 < \epsilon < \min\{\frac{\Delta}{8}, 1 - \left(1 - \frac{\Delta}{8}\right)^{\frac{1}{2N}}, \frac{\tilde{\Delta}}{8(r+p)}\}$.

Step 1: After $\tilde{t}_{\epsilon,\Delta}$ periods in a group the intended action of all responders corresponds with probability of at least $1 - \frac{\Delta}{8}$ to their respective preference type and will not change afterwards.

To see this consider a responder. Once he has tried all actions available at a certain node his valuation of each action will not change further. At each node he will afterwards choose the action which gives him the highest subjective utility with probability $(1-\epsilon)$), i.e. a rewarder will reward after cooperation, a punisher will punish after defection and a self-interested type will neither reward nor punish. Note that after three periods all (of the maximally two) decision nodes of Player 2 have been reached at least once and there is maximally one decision node left that has not yet been reached twice. Let $\tilde{t}_{\epsilon,\Delta}$ be a time with $\tilde{t}_{\epsilon,\Delta} > \frac{ln\left(1-(1-\Delta/8)^{\frac{1}{N}}\right)}{ln(1-\epsilon)} + 3$ For all times $t > \tilde{t}_{\epsilon,\Delta}$ the probability that for any given player all final nodes have been reached (and therefore every action has been tried by Player 2 and his intended action will now always correspond to his preference type) is greater than $1-(1-\epsilon)^{t-3} > \left(1-\frac{\Delta}{8}\right)^{\frac{1}{N}}$. Hence, the probability that all responders in a given group intend to play according to their respective preference type after time $\tilde{t}_{\epsilon,\Delta}$ is greater than $\left(1-\frac{\Delta}{8}\right)$.

Step 2: We show that there exists a time $\tilde{\tilde{t}}_{\epsilon,\Delta}$ such that for all $t > \tilde{t}_{\epsilon,\Delta} + \tilde{\tilde{t}}_{\epsilon,\Delta}$ the probability that all players of any group have an intended action which corresponds to the action derived in Section I[34] is greater than $1-\frac{\Delta}{2}$.

Conditional on being in a group in which the intended actions of all responders correspond to their respective preference types the expected payoff of cooperation and of defection for a proposer is respectively

$$\begin{aligned} E_\epsilon(C) &= c_1 + \frac{k_+}{N_2}r + \epsilon\left(1-\frac{2k_+}{N_2}\right)r \\ E_\epsilon(D) &= d_1 - \frac{k_-}{N_2}p - \epsilon\left(1-\frac{2k_-}{N_2}\right)p \end{aligned} \tag{B7}$$

Since $\epsilon < \frac{\tilde{\Delta}}{8(r+p)}$, we have $|E(C)-E_\epsilon(C)| < \frac{\tilde{\Delta}}{8}$ and $|E(D)-E_\epsilon(D)| < \frac{\tilde{\Delta}}{8}$.

We consider now without loss of generality the case when $E(C) > E(D)$. (The argument for the case $E(C) < E(D)$ is completely analogous.) Then we know from the definition of $\tilde{\Delta}$ that $E(C)-E(D) \geq \tilde{\Delta}$.

By the weak law of large numbers it holds that when the proposer played C sufficiently often that $Prob(|V_t(C)-E_\epsilon(C)| \leq \frac{\tilde{\Delta}}{4}) > \left(1-\frac{\Delta}{8}\right)^{\frac{1}{2N}}$ and when he played D sufficiently often $Prob(|V_t(D)-E_\epsilon(D)| \leq \frac{\tilde{\Delta}}{4}) > \left(1-\frac{\Delta}{8}\right)^{\frac{1}{2N}}$. Whatever this numbers we can choose $\tilde{\tilde{t}}_{\epsilon,\Delta}$ sufficiently large such that the probability that both C and D are played this often is larger than $1-\frac{\Delta}{8}$.

Conditional on being in such a situation, we have

$$V_t(C) \geq E_\epsilon(C) - \frac{\tilde{\Delta}}{4} > E(C) - \frac{\tilde{\Delta}}{4} - \frac{\tilde{\Delta}}{4} \geq E(D) + \tilde{\Delta} - \frac{\tilde{\Delta}}{2} > E_\epsilon(D) - \frac{\tilde{\Delta}}{4} + \frac{\tilde{\Delta}}{2} \geq V_t(D). \tag{B8}$$

[34] I.e. cooperation for $c_1 - d_1 + \frac{k_+r+k_-p}{N_2} > 0$ and defection for $c_1 - d_1 + \frac{k_+r+k_-p}{N_2} < 0$.

Thus for any $T > T_1 \equiv \frac{\left(\tilde{t}_{\epsilon,\Delta}+\tilde{\tilde{t}}_{\epsilon,\Delta}\right)}{\Delta}$ it holds for any group that for periods $t > \Delta T$ the probability that the intended actions of all players in the group corresponds to the action profile derived under Assumption 2 is greater than $\left(1 - \frac{\Delta}{2}\right)$.

Step 3: Considering only those groups, we have i.i.d trembles with probability ϵ, and in any such period we have with a probability of at least $(1-\epsilon)^{2N} > 1 - \frac{\Delta}{8}$ that $a^t_{\mathbf{k}} = a^*_{\mathbf{k}}$. By the law of large numbers we can find a sufficiently large $T_2 > T_1$ such that for all $T > T_2$ with probability greater than $1 - \frac{\Delta}{4}$ from these periods a proportion greater or equal

Together, for $T > T_1$ it holds that for time-periods $t > \Delta T$ we have $W^t_{\mathbf{k}}(a^t_{\mathbf{k}} = a^*_{\mathbf{k}}) \geq (1 - \Delta)$. This establishes that the ϵ-reinforcement learning process is a Δ-bounded perturbation of the main model in Section I.

Within group evolution, formal part

In this section we analyze the robustness of our results with respect to the effect of within-group evolution. One way to think of the assumption that the evolution of preferences occurs only during reshuffling is that players produce offspring in every period which are first in a separate pool of not yet reproducing potential future players. After T periods before groups are reshuffled a small fraction f of the players dies and is replaced randomly by players from the pool of offspring. From this new population new random groups are formed. Then the group composition is always multinomially distributed.

What happens now if we assume instead that evolution can take place in every subperiod. We want to allow for some within-group evolution, while preserving the effect that groups with higher average fitness have more offspring. A natural way to capture this would be to allow the cooperative groups to grow in size relative to the non-cooperative groups. However, given this is a robustness check of our previous model, we want to keep the group size constant at $2N$ interacting players. We therefore allow groups with higher average fitness in their group to have more offspring, but these offspring remain inactive until they either replace an active player or until the next reshuffling of groups occurs.

For any of the three models considered so far, the main model of Section I, the model with trembles, and the model with learning by valuation we consider the following extension which allows for within-group evolutionary forces. In any given subperiod every player passes away with probability $w > 0$, independent across players and subperiods, and is replaced by one of the offspring. Then the expected number of subperiods a player lives is given by $\frac{1}{w}$. Thus, a smaller death probability w corresponds to a larger expected time-span before a player dies and is replaced. We consider situations where the total population does not grow or shrink. This implies that with a lower probability of dying, the overall number of offspring must be smaller, which corresponds to a rescaling of the fitness payoffs by a common factor. Then a small w corresponds to a slow speed of evolution. The precise way that replacements occur turns out to be irrelevant for our robustness check. For concreteness, assume that dead players are replaced with offspring from the same group, if there are any, and with random players from the remaining groups otherwise. In addition, we assume that for any type θ the proportion this type has in the offspring population in the next period is bounded by $C\gamma_\theta$, where $C > 0$ is a potentially large, but finite, constant and γ_θ the proportion of type θ players in the original total population. This assumption is very natural, in particular if we think of situations where every player reproduces at some rate that is independent from the interaction we investigate, plus a growth rate that is derived from the performance in the interaction that we consider.

The key observation for our next result is that if evolution is sufficiently slow relative to the frequency of group reshuffling, than only an arbitrarily small proportion of groups will change their composition and we have a Δ-bounded perturbation of the main model.

LEMMA 5: *Fix a setting. Let $\Delta > 0$.*

(a) Consider the main model of Section I with the above extension which allows for in-group evolutionary forces. For any T there exists a $\overline{w} > 0$ such that for all probabilities of replacement $w < \overline{w}$ this extended model is a Δ-bounded perturbation of the main model.

(b) Consider the model with trembles with the above extension which allows for within-group evolutionary forces. For any T there exist an $\epsilon > 0$ and a $\overline{w} > 0$ such that if $\varepsilon^t_{i\theta}(\mathbf{k}) < \epsilon$ for all $i \in \{1, 2\}$, $\theta \in \Theta$, $t \in \{1, \ldots, T\}$, and $\mathbf{k}$, then, for all probabilities of replacement $w < \overline{w}$, this extended model with errors bounded by ϵ is a Δ-bounded perturbation of the main model in Section I.

(c) Consider the learning by valuation model with the above extension which allows for within-group evolutionary forces. For sufficiently small $\epsilon > 0$ there exists a $T \in \mathbb{N}$ such that for all $T > T_{\epsilon,\Delta}$ there is a $\overline{w}_T$ such that for all $w < \overline{w}_T$ the ϵ-reinforcement learning process by valuation (which replaces Assumption 3 of an observable group composition) extended by above extension allowing for within-group evolution is a Δ-bounded perturbation of the main model in Section I.

B7. Proof of Lemma 5:

We only have to show that $\left|P^t_{\theta,\mathbf{k}}(\boldsymbol{\gamma}) - B_{\theta,\mathbf{k}}(\boldsymbol{\gamma})\right| < \Delta$, holds for sufficiently small $w > 0$ for all $\boldsymbol{\gamma} \in \Gamma$. The property that $W^t_{\mathbf{k}}(a^t_{\mathbf{k}} = a^*_{\mathbf{k}}) \geq (1 - \Delta)$ for all $t > T\Delta$, conditionally on type θ being in a group that has composition $\mathbf{k}$ in all T periods, follows immediately from the previous analysis. (In case (a) we have $a^t_{\mathbf{k}} = a^*_{\mathbf{k}}$ in all periods anyway, in case (b) the proof of Lemma 3 needs only to be conditioned on type θ being in a group that has composition $\mathbf{k}$ in all T periods, and in case (c) the same is true for the proof of Lemma 4.)

Let $f^t_{\theta,w}(\boldsymbol{\gamma})$ denote the fraction of type θ players in subperiod t which are born after the last reshuffling. Then, with $P^T_{\mathbf{k}}(w)$ denoting the probability that a given group of composition $\mathbf{k}$ has never a change in its composition in all T periods, for sufficiently small w holds

$$B_{\theta,\mathbf{k}}(\boldsymbol{\gamma}) \geq P^t_{\theta,\mathbf{k}}(\boldsymbol{\gamma}) \quad \geq \quad (1-f^t_{\theta,w}(\boldsymbol{\gamma}))P^T_{\mathbf{k}}(\mathbf{w})B_{\theta,\mathbf{k}}(\boldsymbol{\gamma}) \geq (1-\frac{\Delta}{2})(1-\frac{\Delta}{2})B_{\theta,\mathbf{k}}(\boldsymbol{\gamma}) \geq B_{\theta,\mathbf{k}}(\boldsymbol{\gamma})-\Delta.$$

This holds since, when w goes to zero, $P^T_{\mathbf{k}}(w) = (1-w)^{NT}$ and $(1-f^t_{\theta,\boldsymbol{\gamma}}(w))$ both converge uniformly to one. This follows since the proportion of any type θ in the offspring population is bounded from above by $\gamma_\theta C$, where $C>0$ is a positive constant which does not depend on $\boldsymbol{\gamma}$. Thus, $f^t_{\theta,w}(\boldsymbol{\gamma}) \geq \frac{\gamma_\theta C\left(1-(1-w)^t\right)}{\gamma_{\theta(1-w)^t}} = C\frac{1-(1-w)^t}{(1-w)^t}$, where $(1-w)^t$ is the proportion of players (of any type in the total population) in period t already born before period 1, and $\left(1-(1-w)^t\right)$ is the proportion of offsprings born after the last group reshuffling.

Robustness Checks on Endogenous Reward and Punishment Intensities

Consider, e.g. a modified version of Setting 1 where the responder can choose a pair $\{r, c(r)\}$ from the available rewarding technologies, where $c(r)$ is the most cost efficient way to achieve reward level r, $c(r)$ is strictly increasing and continuous in r, and $c(0)=0$. The responder chooses a reward level $r \in [0, R_{max}]$ corresponding to his preference type.

First we consider which of the previously derived equilibria are stable against local mutations where the reward chosen by such a mutant is only slightly different compared to the existing types. For reward levels for which there is a k in $\{1,...,N-1\}$ such that $r_k = (d_1-c_1)N/k$, we have cooperation in groups with k or more r_k-level rewarders. A mutant with marginally smaller $\tilde{r}$ can not invade successfully since, in first order, he is the only $\tilde{r}$ rewarder in his group and is thus never pivotal. A marginally larger $\tilde{r}$ cannot invade successfully either, since he has higher costs, but only the same chance of being pivotal as the r_k rewarders. Hence, there are N-1 equilibria which are stable against local mutations and cannot be eroded by gradual shifts in the reward level.

Which of these potential (N-1) evolutionary stable equilibria is actually stable against all mutations seems harder to analyze in general, but for specific parameters and cost functions it is straightforward to check which of these N-1 candidates is actually stable. Suppose the rewarders in the candidate equilibrium choose the reward level r_k. Essentially, for each candidate equilibrium, there are only k mutant types to check (the one who plays an r just high enough to induce cooperation in a group with (k-1) r_k-type rewarders, the one with r' just enough to induce cooperation with k-2 r_k-type rewarders,. . . .)

Note that the equilibrium without any rewarders is not stable against invasion of any rewarder type who could induce cooperation, while e.g. the equilibrium at the reward level $r_1 = (d_1 - c_1)N$ (i.e. k=1, which means a single rewarder is enough to induce cooperation) is definitely stable. Which further candidate equilibria with rewarders are also stable will depend on the parameters.

B8. Comparative Statics of γ_-^{cut} in Setting 2

Let γ_-^{cut} be the fraction of punishers in the unstable mixed equilibrium. This fraction separates the basins of attraction of the stable equilibria. If the initial fraction of punishers is below the cutoff γ^{cut} then this fraction decreases until the entire population has self-interested preferences and nobody punishes defection. If, on the other hand, the initial fraction of punishing players is above the cutoff γ_-^{cut}, then this fraction increases until the entire population has preferences for punishing. One might therefore interpret the value of γ_-^{cut} as an indicator for how likely it is to end up in one or the other equilibrium. The comparative statics of γ_-^{cut} is analogous to Setting 1 and can be directly derived from Equation A5.

Higher costs of punishing diminish the basin of attraction of the punisher equilibrium:

PROPOSITION 6: *If the costs c_p - which a Player 2 has to bear in order to punish - increase, then γ_-^{cut} increases, i.e. there have to be initially more punishers in order to end up in the punishing equilibrium. Furthermore,* $\lim_{c_p \to 0} \gamma_-^{cut} = 0$ *and* $\lim_{c_p \to \infty} \gamma_-^{cut} = 1$.

The intuition is straightforward: the higher the number of punishing players, the cheaper it is to be a punisher. If the costs of punishing increase, punishers become less fit. Hence punishing players need a higher fraction of punishers in order to be at least as successful as non-punishers.

Higher gains from cooperation for Player 2 are good for punishers. Hence the basin of attraction for their equilibrium becomes larger:

PROPOSITION 7: *If the gains of cooperation for the players 2 $(c_2 - d_2)$ increase, then γ_-^{cut} decreases, i.e. a lower initial fraction of punishing players is necessary in order to end up in the punishing equilibrium. Furthermore,* $\lim_{(c_2-d_2)\to 0} \gamma_-^{cut} = 1$ *and* $\lim_{(c_2-d_2)\to \infty} \gamma_-^{cut} = 0$.

Again, the intuition is straightforward: the higher the gains of cooperation for a Player 2, the higher his profit from being pivotal in inducing cooperation of players 1. Therefore a lower fraction of punishers is necessary in order to make punishing more successful than non-punishing.

LEMMA 6: *If the threshold k_-^* of punishing players 2 in a group above which the players 1 start to cooperate increases then γ_-^{cut} increases, i.e. there are more punishing players necessary in order to end up in the punishing equilibrium.*

Intuitively, a higher threshold k^*_- makes it more probable for an individual to be in a group in which the number of punishers is too low to induce cooperation. In these groups being a punisher is costly. Therefore, fitness of punishers is lower and a higher initial fraction of punishers is necessary to make punishing more successful than non-punishing.

COROLLARY 2: *If Player 1's costs for cooperation* $(d_1 - c_1)$ *increase, then* γ_-^{cut} *increases weakly, i.e. a weakly higher punisher-proportion is necessary in order to end up in the punishing equilibrium.*

COROLLARY 3: *If Player 2's losses due to a punishment* p *increase, then* γ_-^{cut} *decreases weakly.*

B9. *Definitions from Evolutionary Game Theory*

The following presentation adapts definitions from Weibull (1995), Chapter 6, to our setting.

DEFINITION 3: *A regular growth-rate function is a Lipschitz-continuous function* $\mathbf{g} : \mathcal{X} \to \mathbb{R}^{|\Theta|}$ *with open domain* $\mathcal{X} \subset \mathbb{R}^{|\Theta|}$ *containing* Γ*, such that* $\mathbf{g}(\boldsymbol{\gamma}) \cdot \boldsymbol{\gamma} = 0$ *for all* $\boldsymbol{\gamma} \in \Gamma$*.*

These regularity conditions on the growth rates guarantee a unique solution to the system of differential equations 2, as given earlier. The condition $\mathbf{g}(\boldsymbol{\gamma}) \cdot \boldsymbol{\gamma} = 0$ guarantees that the sum of the population shares remains one, since $\sum_{\theta \in \Theta} \dot{\gamma}_\theta = \mathbf{g}(\boldsymbol{\gamma}) \cdot \boldsymbol{\gamma} = 0$ *for all* $\boldsymbol{\gamma} \in \Gamma$. For any initial state $\boldsymbol{\gamma}^{\mathbf{0}} \in \Gamma$, the dynamic will remain in Γ. Together with the Lipschitz continuity on an open domain $\mathcal{X} \supset \Gamma$, this guarantees a unique solution $\boldsymbol{\xi}(\cdot, \boldsymbol{\gamma}^{\mathbf{0}}) : \mathbb{R} \to \mathcal{X}$ through every initial state $\boldsymbol{\gamma}^{\mathbf{0}} \in \Gamma$. Furthermore, $\boldsymbol{\xi}(\tau, \boldsymbol{\gamma}^{\mathbf{0}})$ is continuous in $\tau \in \mathbb{R}$ and $\boldsymbol{\gamma}^{\mathbf{0}} \in \Gamma$. Definition 4 relates the evolutionary success to the fitness obtained. Payoff monotonicity captures the process of selection: Types that earn greater average fitness have higher growth rates.

DEFINITION 4: *A regular growth-rate function* $\mathbf{g}$ *is called payoff-monotonic if:*
$\bar{\pi}_\theta(\boldsymbol{\gamma}) < \bar{\pi}_{\theta'}(\boldsymbol{\gamma}) \Leftrightarrow g_\theta(\boldsymbol{\gamma}) < g_{\theta'}(\boldsymbol{\gamma}) \qquad$ *for all* $\theta, \theta' \in \Theta, \boldsymbol{\gamma} \in \Gamma$*.*

DEFINITION 5: *A population state* $\boldsymbol{\gamma} \in \Gamma$ *is called* **Lyapunov-stable** *if every neighborhood* $\mathcal{B}$ *of* $\boldsymbol{\gamma}$ *contains a neighborhood* $\mathcal{B}^0$ *of* $\boldsymbol{\gamma}$ *such that* $\boldsymbol{\xi}(\tau, \boldsymbol{\gamma}^{\mathbf{0}}) \in \mathcal{B}$ *for all* $\boldsymbol{\gamma}^{\mathbf{0}} \in \mathcal{B}^0 \cap \Gamma$ *and for all* $\tau \geq 0$*.*

A state $\boldsymbol{\gamma} \in \Gamma$ *is called* **(asymptotically) stable** *if it is Lyapunov-stable and there exists a neighborhood* $\mathcal{B}^*$ *such that the equation* $\lim_{\tau \to \infty} \boldsymbol{\xi}(\tau, |bf\gamma^0) = \boldsymbol{\gamma}$ *holds for all* $\boldsymbol{\gamma}^{\mathbf{0}} \in \mathcal{B}^* \cap \Gamma$*.*

Whether a limit point of the dynamics is robust to mutations is captured by the stability of this state. Intuitively, Lyapunov stability guarantees that sufficiently small mutations do not lead far away from the population state under consideration. Asymptotic stability requires in addition that after small mutations the population is driven back to the stable state by the selection dynamics. These stability concepts can be generalized to sets.

DEFINITION 6: *A* **closed set** $\mathcal{A} \subset \Gamma$ *is* **Lyaponov-stable** *if every neighborhood* $\mathcal{B}$ *of* $\mathcal{A}$ *contains a neighborhood* $\mathcal{B}^0$ *of* $\mathcal{A}$ *such that* $\boldsymbol{\xi}(\tau, \boldsymbol{\gamma^0}) \in \mathcal{B}$ *for all* $\boldsymbol{\gamma^0} \in \mathcal{B}^0 \cap \Gamma$ *and for all* $\tau \geq 0$.

A closed set $\mathcal{A} \subset \Gamma$ *is* **asymptotically stable** *if it is Lyapunov-stable and if there exists a neighborhood* $\mathcal{B}^*$ *of* $\mathcal{A}$ *such that* $\boldsymbol{\xi}(\tau, \mathbf{x^0})_{\tau \to \infty}$ *converges to* $\mathcal{A}$ *for all* $\mathbf{x^0} \in \mathcal{B}^* \cap \Gamma$. [35]

[35] We say $\boldsymbol{\xi}(\tau, x^0)_{\tau \to \infty}$ *converges to* $\mathcal{A}$, where $\mathcal{A}$ is a closed set, if the distance $d(\boldsymbol{\xi}(\tau, x^0), \mathcal{A})\tau \to \infty$ *converges to* 0. The distance $d(\mathbf{y}, \mathcal{A})$ between a point $\mathbf{y} \in \Gamma$ and a closed set $\mathcal{A}$ is defined as the minimal distance between $\mathbf{y}$ and any point $\mathbf{a} \in \mathcal{A}$, i.e., $d(\mathbf{y}, \mathcal{A}) = \min_{\mathbf{a} \in \mathcal{A}} d(\mathbf{y}, \mathcal{A})$.