1. Model

We consider a mixed population of two types, selfish and group reciprocators (GR). At the beginning of each generation, the population randomly divides into a large number of groups of size G each. Let p denote the population share of GR types. Let p_g be the proportion of GR in group g, which is distributed binomially.

At every step t, everybody interacts with everybody. In each pair, each individual chooses between cooperation and defection. Cooperation entails a cost c to the cooperator and a benefit b to her partner. Defection carries no costs. That is, each pair plays the following Prisoner's Dilemma game:

	Cooperate	Defect
Cooperate	$b-c$	$-c$
Defect	b	0

Selfish types always defect, A GR individual i starts by cooperating, and then cooperates with all individuals belonging to group g with a probability $\phi(l_{gi})$, where l_{gi} is the proportion of individuals from group g who cooperated with individual i in round $t-1$. $\phi(\cdot)$ is monotonically weakly increasing. We consider the cutoff strategy:

$$\phi(l_{gi}) = \begin{cases} 1 & \text{if } l_{gi} \geq k \\ 0 & \text{otherwise.} \end{cases}$$

The fitness is the payoff at the limit where $t \to \infty$. Equivalently, since the game always settles to a stationary action profile, it is the average payoff of T rounds when $T \to \infty$.

Given this behaviour, GR in all groups where $p_g \geq k$ help all individuals in other groups where $p_g \geq k$ and defect against members of all other groups. GR in other groups always defect. Individuals' fitness therefore depends only on whether they are in a "supraliminal" group with $p_g \geq k$, and on their type. Let q be the proportion of supraliminal groups. Let $\bar{p}$ be the proportion of GR individuals in supraliminal groups (out of the total population in such groups).

Let $\underline{p}$ be the proportion of GR individuals in subliminal groups (out of the total population in such groups). It follows that

- Group reciprocators in supraliminal groups get a payoff of $\bar{p}qb - qc$.
- Selfish types in supraliminal groups get $\bar{p}qb$.
- Group reciprocators and selfish types in subliminal groups get 0.

After each generation, reproductive success is proportional to fitness, the total population size stays the same, and children are remixed randomly into new groups of the same size. The mean fitness of the GR type is

$$\frac{\bar{p}q\left(q\left(\bar{p}b - c\right)\right)}{p}$$

and the mean fitness of selfish types is

$$\frac{(1-\bar{p})q\left(q\bar{p}b\right)}{1-p}$$

After rearranging, the mean fitness of reciprocators is higher if

$$\frac{\bar{p}-p}{1-p} \geq \frac{c}{b}. \tag{1}$$

where

$$\bar{p} = E[p_g|p_g \geq k] = \frac{1}{G}\frac{\sum_{l=kG}^{G} l\text{Binom}(l, G, p)}{\sum_{l=kG}^{G} \text{Binom}(l, G, p)}$$

The LHS of (1) is decreasing in p and is equal to the threshold k when $p = 0$.[1] It follows that there is a unique ESS with a positive share of group reciprocators if and only if $k > \frac{c}{b}$. Otherwise the population is homogeneously selfish in the unique ESS.

[1] When the share of GR in the population is very small, the probability that $p_g = k$ conditional on $p_g \geq k$ goes to one.