Supporting information

1. Inclusive fitness of a focal Role 2 actor

In the main text, equations (2)-(5) show how we set up the inclusive fitness function for a randomly chosen focal actor in Role 1 (w_1). Here, we give the equations used to set up the corresponding function w_2 for a randomly chosen focal actor in Role 2.

The focal Role 2 actor plays x_2 in a group of individuals playing x'_2 and x'_1. Each of these types of individuals acquires the following respective fractions of the group's resources:

$$q_2 = \frac{bx_2}{b(x_2 + (n_2 - 1)x'_2) + n_1 x'_1} \tag{S1a}$$

$$q'_2 = \frac{bx'_2}{b(x_2 + (n_2 - 1)x'_2) + n_1 x'_1} \tag{S1b}$$

$$q'_1 = \frac{x'_1}{b(x_2 + (n_2 - 1)x'_2) + n_1 x'_1} \tag{S1c}$$

The total cooperative effort of the focal individual's group is:

$$c_2 = (1 - x_2) + (n_2 - 1)(1 - x'_2) + n_1(1 - x'_1) \tag{S2}$$

The fraction of resource obtained by this group (i.e. the group productivity) is as follows (c' is given in equation 3b in the main text):

$$p_2 = \frac{c_2}{c_2 + (g - 1)c'} \tag{S3}$$

Using equation 1 in the main text, we obtain the inclusive fitness of the focal Role 2 actor:

$$w_2 = \frac{\partial v p_2 q_2}{\partial x_2} + r(n_2 - 1)\frac{\partial v p_2 q_2'}{\partial x_2} + r n_1 \frac{\partial v p_2 q_1'}{\partial x_2} \quad \text{(S4)}$$

2. Neighbor-modulated fitness and definitions of relatedness

In the inclusive fitness model in the main text, we considered the effect of a slight change in the selfish strategy x of a randomly chosen focal actor on its group-mates playing strategy x', following Taylor *et al.* 's (2007) equation 19. We defined r as the weight the focal actor gives to the fitness increments of each member of its group (Hamilton, 1964). Here, we present a neighbor-modulated or direct fitness model. This approach considers the alleles underlying the selfish strategies x'_1 and x'_2, tracking the effect of a mutation in each of these alleles (x_1 and x_2 respectively) on a randomly chosen focal recipient, and gives the same results as the inclusive fitness model in the main text (Taylor *et al.*, 2007).

In the following sections, we (a) show how the neighbor-modulated fitness function is constructed; (b) prove that the definition of relatedness that we use in the neighbor-modulated fitness model (the probability that a group member has the mutation, given that a focal individual has it) is the same as the relatedness coefficient that arises from the fitness function as the regression of the partner's breeding value on the focal's breeding value (Wenseleers *et al.*, 2004, 2010; Gardner *et al.*, 2011); and (c) demonstrate that this also yields the definition of relatedness that we use in the inclusive fitness model (the weight given by the focal actor to its effect on each other individual's fitness (Hamilton, 1964)).

a. Neighbor-modulated fitness function

We assume the selfish efforts of individuals in Role 1 and Role 2 are governed by alleles x'_1 and x'_2 respectively, and that each allele is at a separate locus. (If we instead assume that there is a single locus, with x'_1 found only in Role 1 and x'_2 only in Role 2, we obtain the same solutions; results not shown.) We consider a mutation in each allele: x_1 and x_2 respectively. We define r as the probability that the mutation is found in another group member, given that one individual has it. We assume that individuals are haploid and that mutations are rare, such that no individuals have mutations at both loci.

We first consider the mutant allele for Role 1 behavior (x_1). If the mutation is found in a Role 1 player, it plays x_1, but if a Role 2 player has the mutation, it plays the non-mutant strategy x'_2. The fractions of resource obtained by a focal Role 1 player and a Role 2 player each carrying the mutant Role 1 allele are:

$$q_{11NMF} = \frac{x_1}{x_1 + (n_1 - 1)(rx_1 + (1-r)x'_1) + n_2 b x'_2} \tag{S5a}$$

$$q_{12NMF} = \frac{b x'_2}{n_1(rx_1 + (1-r)x'_1) + n_2 b x'_2} \tag{S5b}$$

Below are the total cooperative efforts of a group where x_1 is in Role 1 and a group where x_1 is in Role 2 respectively:

$$c_{11NMF} = (1 - x_1) + \ (n_1 - 1)(r(1 - x_1) + (1-r)(1 - x'_1)) + n_2(1 - x'_2) \tag{S6a}$$

$$c_{12NMF} = n_1(r(1 - x_1) + (1-r)(1 - x'_1)) + n_2(1 - x'_2) \tag{S6b}$$

Using *c'* from equation 3b in the main text, we find the group shares:

$$p_{11NMF} = \frac{c_{11NMF}}{c_{11NMF} + (g-1)c'} \tag{S7a}$$

$$p_{12NMF} = \frac{c_{12NMF}}{c_{12NMF} + (g-1)c'} \tag{S7b}$$

The mutant allele x_1 is in a Role 1 player with probability $\frac{n_1}{n_1+n_2}$, and in a Role 2 player with probability $\frac{n_2}{n_1+n_2}$. The expected direct fitness increment of an individual carrying this mutation (see Taylor *et al.*'s (2007) equation 17) is thus:

$$w_{1NMF} = \frac{n_1}{n_1+n_2}\frac{\partial v p_{11NMF} q_{11NMF}}{\partial x_1} + \frac{n_2}{n_1+n_2}\frac{\partial v p_{12NMF} q_{12NMF}}{\partial x_1} \tag{S8}$$

We now follow the same procedure to obtain an expression for w_{2NMF}, the direct fitness increment of an individual carrying the mutant allele for the Role 2 strategy, x_2. In this case, a Role 2 player with the mutation plays x_2, and a Role 1 player with the mutation plays x'_1.

Individual shares of resource when x_2 is found in a Role 1 and Role 2 player respectively:

$$q_{21NMF} = \frac{x'_1}{n_1 x'_1 + n_2 b(rx_2 + (1-r)x'_2)} \tag{S9a}$$

$$q_{22NMF} = \frac{b x_2}{n_1 x'_1 + b(x_2 + (n_2 - 1)(rx_2 + (1-r)x'_2))} \tag{S9b}$$

Group cooperative efforts:

$$c_{21NMF} = n_1(1 - x_1') + n_2(r(1 - x_2) + (1 - r)(1 - x_2')) \tag{S10a}$$

$$c_{22NMF} = n_1(1 - x_1') + (1 - x_2) + (n_2 - 1)(\, r(1 - x_2) + (1 - r)(1 - x_2')) \tag{S10b}$$

Group shares:

$$p_{21NMF} = \frac{c_{21NMF}}{c_{21NMF} + (g - 1)c'} \tag{S11a}$$

$$p_{12NMF} = \frac{c_{22NMF}}{c_{22NMF} + (g - 1)c'} \tag{S11b}$$

Expected direct fitness increment of an individual carrying the mutant allele x_2:

$$w_{2NMF} = \frac{n_1}{n_1+n_2}\frac{\partial v p_{21NMF} q_{21NMF}}{\partial x_2} + \frac{n_2}{n_1+n_2}\frac{\partial v p_{22NMF} q_{22NMF}}{\partial x_2} \tag{S12}$$

We then simultaneously solve w_{1NMF}=0 and w_{2NMF} =0 for the optimal values x_1* and x_2* respectively, setting $x_1 = x'_1 = x_1$* and $x_2 = x'_2 = x_2$*, and checking that the derivatives ($\frac{\partial w_{1NMF}}{\partial x_1}$ and $\frac{\partial w_{2NMF}}{\partial x_2}$) are negative. The solutions for x_1* and x_2* were identical to those we got from the inclusive fitness functions, as shown in the attached Mathematica file.

b. Relatedness in the neighbor-modulated fitness model

Let the neighbor-modulated (direct) fitness be expressed as the following function:

$$f[x, r(n - 1)x + (1 - r)(n - 1)x'] \tag{S13}$$

where the first argument refers to the selfish effort of a randomly chosen focal recipient carrying the mutant allele *x*, and the second argument refers to the summed efforts of all other *n*-1 interactants in the group, each of which has a probability *r* of also possessing the mutant allele *x*, given that the focal recipient has it.

The average neighbor modulated fitness of a member of the focal individual's group is:

$$\frac{r(n - 1)f[x, r(n - 1)x + (1 - r)(n - 1)x'] + (1 - r)(n - 1)f[x', r(n - 1)x \; + \; (1 - r)(n - 1)x']}{(n - 1)} \tag{S14}$$

This simplifies to:

$$rf[x, r(n-1)x + (1-r)(n-1)x'] + (1-r)f[x', r(n-1)x \; + \; (1-r)(n-1)x'] \tag{S15}$$

The second arguments of the functions *f* in (S15) are the same, so we can simplify the notation by letting the focal's fitness be equal to *f* and the fitness of any group member that does not carry the mutant allele be equal to *f'*, the latter being taken as fixed and a constant for the population.

The breeding value is the average effect of the allele on the phenotype in offspring. Let *h* be the effect of possession of the rare mutation causing selfish effort *x*, where *h* = *x*-*x'*. As we are assuming haploidy, all of the focal's offspring inherit the mutation, so the focal's breeding value is *h*. (Note that if individuals were diploid, the focal's breeding value would instead be $\frac{h}{2}$. As we would then regress $\frac{hr}{2}$ on $\frac{h}{2}$ instead of *hr* on *h* (see below), we obtain the same answer as for diploid individuals.) The expected breeding value of another member of the group is as follows. (Since breeding value is the average effect in offspring, this expression takes into account the number of offspring, which differs between partners who have the mutation and those who do not.)

$$\frac{r(n-1)fh \; + \; (1-r)(n-1)f'.0}{r(n-1)f \; + \; (1-r)(n-1)f'} \tag{S16}$$

We want to regress the breeding value of another group member on the breeding value of the focal, *h*, across different values of *h*. At equilibrium, *f* = *f'*, so the above expression for a group member's breeding value simplifies to *hr*. We therefore regress *hr* on *h* for different values of *h*.

This gives a regression coefficient of:

$$\frac{Cov[hr, h]}{Var[h]} \tag{S17}$$

From the properties of covariance and variance, this yields:

$$\frac{rVar[\mathrm{h}]}{Var[\mathrm{h}]} \tag{S18}$$

This simplifies to *r*.

Thus, relatedness defined as the probability of a group member possessing the mutant allele, given the focal has it, is the same as relatedness given from the fitness function by regressing the expected breeding value of an interactant on the breeding value of the focal, which is the standard definition for relatedness in neighbor-modulated fitness (Wenseleers *et al.*, 2004, 2010; Gardner *et al.*, 2011).

c. Relatedness in the inclusive fitness model

Let the neighbor-modulated fitness be expressed as the function *f* in S13 above: $f[x, r(n-1)x + (1-r)(n-1)x']$. This gives the effect on a randomly chosen focal recipient's fitness of its carrying the mutant allele *x* in a group with relatedness r. Since the mutant allele is rare, *r* is just the proportion of other individuals that also have the mutant allele *x*; all others have the allele *x'* (see section 2b above).

We find the evolutionarily stable solution *x** as normal (see section 2a above and main text) by seeking $x = x' = x^*$ for which $\frac{\partial f}{\partial x}$=0 (Maynard Smith, 1982; Taylor *et al.*, 2007). This yields the solution equation:

$$f^{(1,0)}[x^*, (n-1)x^*] + r(n-1)f^{(0,1)}[x^*, (n-1)x^*] = 0 \tag{S19}$$

where $f^{(1,0)}$ refers to the first derivative of *f* with respect to *x* for the first argument only and $f^{(0,1)}$ refers to the first derivative of *f* with respect to *x* for the second argument only.

Inclusive fitness, which focuses on fitness effects dispensed to a focal actor rather than received by a focal recipient, is given by:

$$I \;=\; f[x, (n-1)x'] \;+\; r(n-1)f[x', x + (n-2)x'] \tag{S20}$$

The first function represents the effects on the focal actor's fitness of its selfish effort *x* (first argument) and its *n*-1 group members' efforts *x'* (second argument). The second function represents the effects on another group member's fitness of its own selfish effort *x'* (first argument) and those of all other individuals in the group (the focal, who invests *x*, and the *n*-2 other group members, each of which invests *x'*). The focal individual weights each group member's fitness by the same *r* that appears in the neighbor-modulated fitness expression.

We find the evolutionarily stable solution *x** as before, by seeking $x = x' = x^*$ for which $\frac{\partial f[x,(n-1)x']}{\partial x} + \; r(n-1)\frac{\partial f[x',x+(n-2)x']}{\partial x}$ = 0. This leads to exactly the same solution equation as that for neighbor-modulated fitness above, showing that the definitions of relatedness are equivalent.

References

Gardner, A., West, S.A. & Wild, G. 2011. The genetical theory of kin selection. *J. Evol. Biol.* **24**: 1020–1043.

Hamilton, W.D. 1964. The genetical evolution of social behaviour. II. *J. Theor. Biol.* **7**: 17–52.

Maynard Smith, J. 1982. *Evolution and the Theory of Games*. Cambridge University Press, Cambridge.

Taylor, P.D., Wild, G. & Gardner, A. 2007. Direct fitness or inclusive fitness: how shall we model kin selection? *J. Evol. Biol.* **20**: 301–309.

Wenseleers, T., Gardner, A. & Foster, K.R. 2010. Social evolution theory: a review of methods and approaches. In: *Social Behaviour: Genes, Ecology and Evolution* (T. Székely, A. J. Moore, & J. Komdeur, eds), pp. 132–158. Cambridge University Press, Cambridge.

Wenseleers, T., Helanterä, H., Hart, A. & Ratnieks, F.L.W. 2004. Worker reproduction and policing in insect societies: an ESS analysis. *J. Evol. Biol.* **17**: 1035–1047.

3. Mathematica code for numerical solutions

Please see the two attached Mathematica files, in which we give the code we used to generate the numerical solutions, and demonstrate that the inclusive fitness and neighbor-modulated fitness methods are equivalent, yielding the same numerical solutions.

4. Figures for different values of *g*

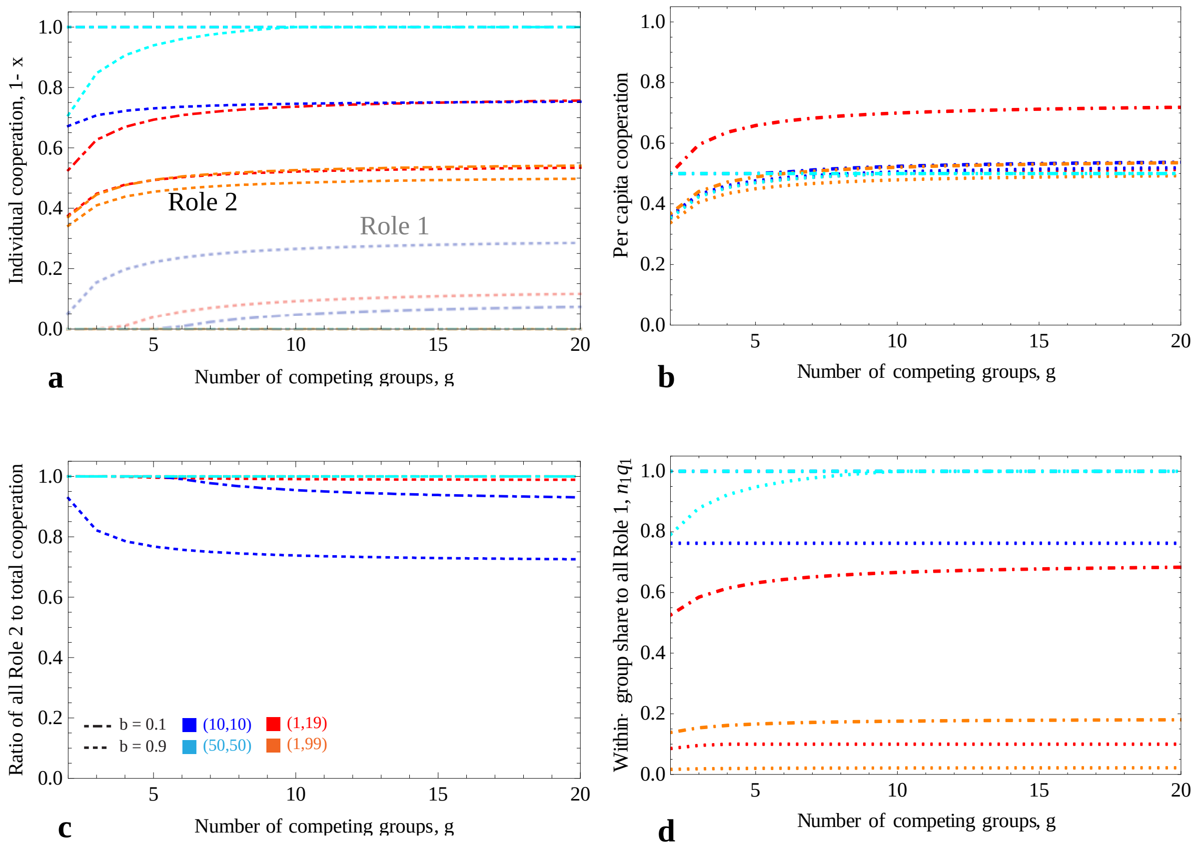
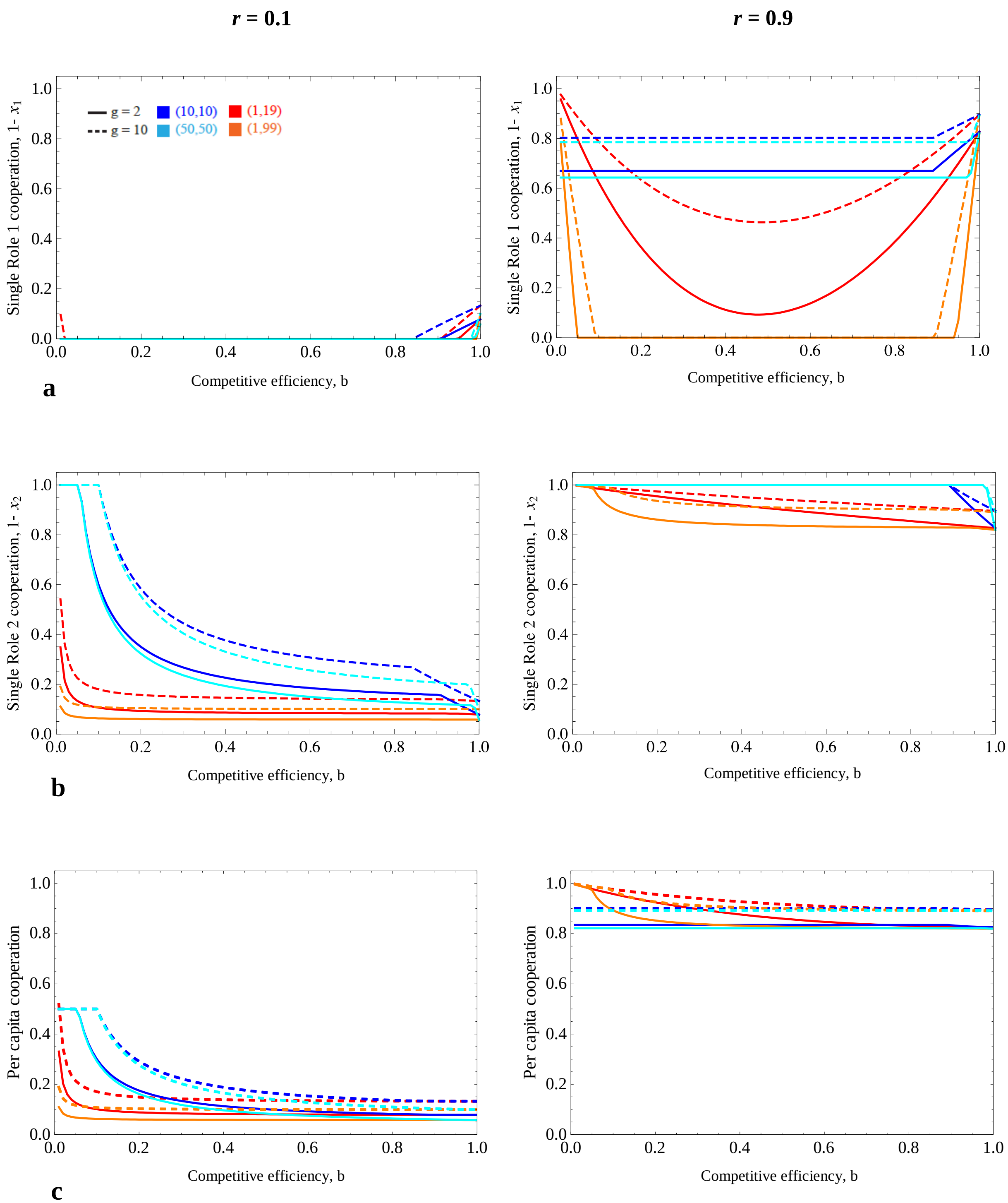
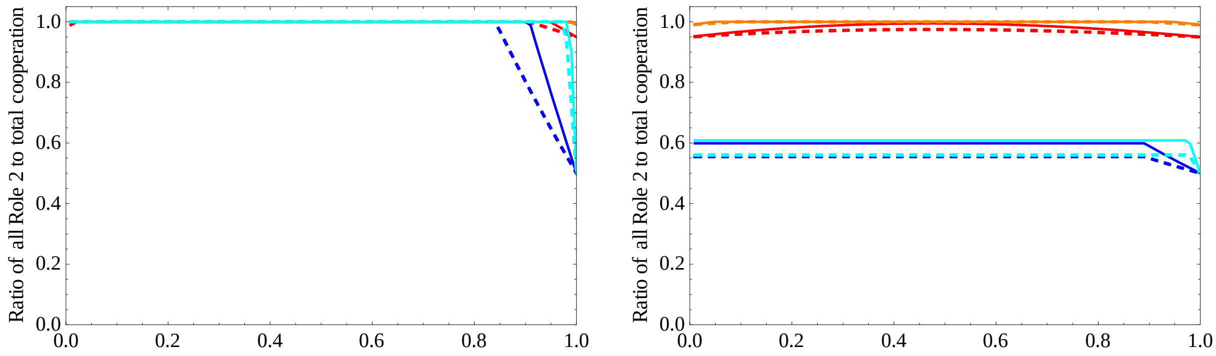


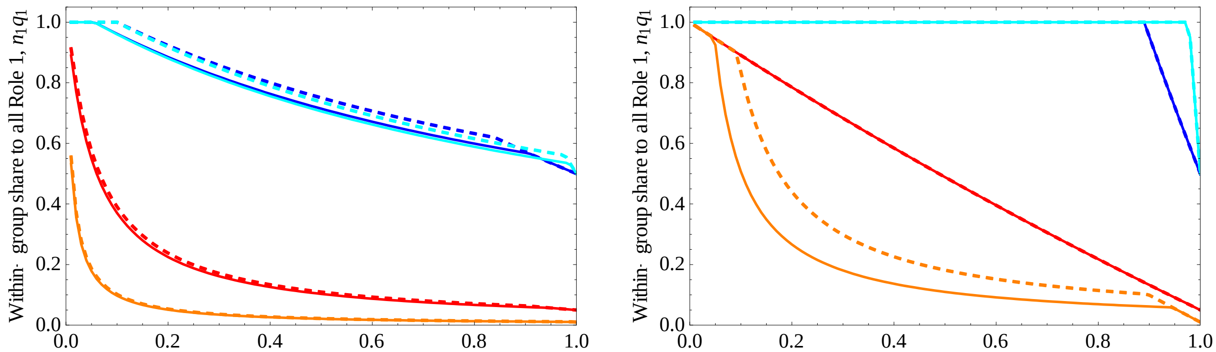
Figure S1. The effect of the number of competing groups, *g*, on: (a) The individual cooperative efforts, 1-x*, of a single individual in Role 1 (faded lines) and Role 2 (darker lines). (b) A group's per capita cooperation: (n_1(1-x_1*)+n_2(1-x_2*)) / (n_1+n_2). (c) The combined cooperative efforts of all Role 2 individuals relative to the group's total cooperation: n_2(1-x_2*) / (n_1(1-x_1*)+n_2(1-x_2*)). (d) The combined share of reproduction obtained by all Role 1 individuals relative to the total reproduction in the group: n_1q_1 / (n_1q_1 + n_2q_2). All panels show solutions for relatedness r=0.5 and resource value v=1, and use the same color scheme. Dot-dashed lines show b=0.1 (high asymmetry in relative competitive efficiency between the two roles) and dotted lines show b=0.9 (low asymmetry in relative competitive efficiency). Blue and cyan lines show groups with equal numbers of Role 1 and Role 2 individuals (n_1=n_2); red and orange lines show groups with a single Role 1 player and many Role 2 players (n_1=1, n_1<<n_2). Darker colors (blue and red) show groups of total size 20 (n_1+n_2), and lighter colors (cyan and orange) show groups of total size 100.

5. Figures for different values of *r*





d



e

Figure S2. The effect of Role 2 relative competitive efficiency, *b*, for different values of relatedness, r. Panels on the left show r=0.1, and panels on the right show r=0.9; Fig. 1 in the main text shows r=0.5. (a) The individual cooperative effort of a single Role 1 player, $1\text{-}x_1$*. (b) The individual cooperative effort of a single Role 2 player, $1\text{-}x_2$*. (c) A group's per capita cooperation: $(n_1(1\text{-}x_1^*)+n_2(1\text{-}x_2^*))$ / (n_1+n_2). (d) The combined cooperative efforts of all Role 2 individuals relative to the group's total cooperation: $n_2(1\text{-}x_2^*)$ / $(n_1(1\text{-}x_1^*)+n_2(1\text{-}x_2^*))$. (e) The combined share of reproduction obtained by all Role 1 individuals relative to the total reproduction in the group: n_1q_1 / $(n_1q_1 + n_2q_2)$. All panels show resource value *v*=1, and use the same color scheme. Solid lines show the number of competing groups *g*=2 and dashed lines show *g*=10. Blue and cyan lines show groups with equal numbers of Role 1 and Role 2 individuals ($n_1=n_2$); red and orange lines show groups with a single Role 1 player and many Role 2 players ($n_1=1$, $n_1<<n_2$). Darker colors (blue and red) show groups of total size 20 (n_1+n_2), and lighter colors (cyan and orange) show groups of total size 100.