*Phil. Trans. R. Soc. B.*

PHILOSOPHICAL TRANSACTIONS B

*Phil. Trans. R. Soc. B.*
doi: 10.1098/rstb.2015.0089.

**Electronic Supplementary Material for publication:**

# Negotiation and appeasement can be more effective drivers of sociality than kin selection

Andrés E. Quiñones[1,2]*, G. Sander van Doorn[1], Ido Pen[1], Franz J. Weissing[1], Michael Taborsky[2]

[1]Theoretical Research in Evolutionary Life Sciences, Groningen Institute for Evolutionary Life Sciences, University of Groningen, 9747 AG Groningen, The Netherlands

[2]Department of Behavioural Ecology, Institute of Ecology and Evolution, University of Bern, CH-3032 Hinterkappelen, Switzerland

*Correspondence to: andresqp@gmail.com

## 1. Mathematical analysis

*Fitness and selection gradients with respect to the behavioral traits*

This appendix outlines a general mathematical argument that highlights the key differences between kin-selected and negotiation strategies. As a starting point, let us consider the evolutionary fate of a mutant genotype that is present at low frequency in a population. The mutant genotype can be expressed in helpers and in breeders, and its success is not only determined by the behaviors of mutant individuals themselves but also by the responses of the resident individuals with whom the mutants interact. Therefore, the fitness of the mutant genotype, quantified by the fitness function $W(a,h,a',h')$, generally depends on four phenotypic traits:

$a$ the amount of aggression exerted by mutants in their role as breeder

$h$ the amount of help provided by mutants in their role as helper

$a'$ the amount of aggression received by mutants in their role as helper

$h'$ the amount of help received by mutants in their role as breeder.

The function $W$ can, in principle, be derived from a class-structured population model that describes how mutant individuals move between the different classes (breeder,

helper and floater) in every generation, depending on the patterns of survival and reproduction in the context of the resident population[1,2]. However, rather than presenting such a derivation for a specific model, we here only make assumptions on the qualitative dependence of fitness on the behavioral traits $a$, $h$, $a'$ and $h'$. In particular, we rely on the generic payoff relationships between two individuals involved in a pay-to-stay interaction to derive inequality constraints on the selection gradients, the derivatives of the fitness function that measure the strength and direction of selection acting on each of the behavioral traits.

The selection gradient with respect to behavioral trait $a$ (i.e., the fitness effect of changing $a$ by one phenotypic unit) is found by calculating

$$w_a = \left.\frac{\partial W(a,h,a',h')}{\partial a}\right|_{\substack{a=a'=\hat{a}\\ h=h'=\hat{h}}} . \tag{S1}$$

Note that the derivative is evaluated at $\hat{a}$ and $\hat{h}$, the levels of aggression and help realized in interactions between a resident helper and a resident breeder. Calculating the selection gradients this way allows us to use a linear approximation to quantify the fitness difference between resident and mutant genotypes,

$$\begin{aligned}\Delta W &= W(a,h,a',h') - W\left(\hat{a},\hat{h},\hat{a},\hat{h}\right)\\ &\approx (a-\hat{a})\,w_a + \left(h-\hat{h}\right)w_h + (a'-\hat{a})\,w_{a'} + \left(h'-\hat{h}\right)w_{h'} ,\end{aligned} \tag{S2}$$

provided that the phenotypic difference between mutant and resident is small. The other selection gradients $w_h$, $w_{a'}$ and $w_{h'}$ that appear in this approximation, are calculated in the same way as $w_a$, except that the derivative is taken with respect to $h$, $a'$ and $h'$, respectively.

In any specific model, the selection gradients are (complicated) functions of the resident phenotypes $\hat{a}$ and $\hat{h}$ for two reasons: first, the selection gradients are evaluated at the resident phenotypes; second, when the mutant is rare, resident-resident interactions determine the frequencies of the different classes of individuals and their reproductive values, which both affect fitness. We do not make these dependencies explicit, because we are merely interested in the relative magnitudes of the selection gradients.

In fact, under the pay-to-stay hypothesis, both receiving and expressing aggression are costly, but the costs to the helper (the receiving party) are higher than the costs to the

breeder. Hence, in general, $w_{a'} < w_a < 0$. To facilitate the interpretation of the further results, we define $w_a = -d$ and $w_{a'} = -d'$, where $d$ and $d'$ represent, respectively, the damage (measured as a marginal fitness costs) incurred by the breeder and helper as a direct consequence of aggressive behavior.

The pay-to-stay hypothesis also posits that providing help is costly to the helper, whereas receiving help is beneficial to the breeder, so that $w_h < 0 < w_{h'}$. Motivated by this general inequality constraint, we substitute $w_h$ by $-c$ and $w_{h'}$ by $b$, and henceforth refer to $c$ anb $b$ as the (marginal) fitness cost and benefit of help, respectively.

*Selection gradients with respect to genetic traits that affect the reaction norms*

The next step in the analysis is to take into account that the behavioral phenotypes (*a, h, a', h'*) depend on genetic traits that define the shape of the reaction norms for aggression and help. Consider, for example a trait $x$ that affects the behavior of the helper, so that the help provided by the helper $h=H(x, a')$ depends on the amount of aggression received, but also on the genotype of the helper, $x$. Then, based on the chain-rule for differentiation, the fitness gradient with respect to $x$ is given by

$$\begin{aligned} w_x &= \left.\frac{\partial W(a,h,a',h')}{\partial h}\right|_{\substack{a=a'=\hat{a}\\ h=h'=\hat{h}}} \left.\frac{\partial h}{\partial x}\right|_{x=\hat{x}} + \left.\frac{\partial W(a,h,a',h')}{\partial a'}\right|_{\substack{a=a'=\hat{a}\\ h=h'=\hat{h}}} \left.\frac{\partial a'}{\partial x}\right|_{x=\hat{x}} \\ &= w_h \left.\frac{\partial h}{\partial x}\right|_{x=\hat{x}} + w_{a'} \left.\frac{\partial a'}{\partial x}\right|_{x=\hat{x}} \\ &= -c \left.\frac{\partial h}{\partial x}\right|_{x=\hat{x}} - d' \left.\frac{\partial a'}{\partial x}\right|_{x=\hat{x}} . \end{aligned} \tag{S3}$$

In this derivation, we took into consideration that the behavioral phenotypes $h$ and $a'$ implicitly depend on the genetic trait $x$ because they are equilibria of the negotiation dynamics

$$\begin{cases} h_{t+1} = H(x, a'_t) , \\ a'_{t+1} = A(\hat{y}, h_t) , \end{cases} \tag{S4}$$

in which $x$ affects the helper reaction norm. Equations (S4) also makes explicit that the breeder's response to the help offered by the helper, $A(y, h)$, may depend on a different genetic trait $y$. At this point, however, we keep $y$ fixed at the resident trait value $\hat{y}$, since individuals carrying a mutation in the helper reaction norm interact with breeders expressing a resident reaction norm for aggression. In other words, mutations are considered to occur one at a time. Finally, we note that the fitness

gradients $w_a$ and $w_{h'}$ do not appear in equation (S3) because these relate to the mutant's performance as a breeder, which is independent of the genetic trait $x$.

The fact that the combination of behavioral traits ($h$, $a'$) is an equilibrium of the negotiation dynamics (S4), gives rise to the following result:

$$\begin{aligned} & h = H\left(x, A\left(\hat{y}, h\right)\right) && \Rightarrow \\ & \left.\frac{\partial h}{\partial x}\right|_{x=\hat{x}} = \left.\frac{\partial H\left(x,\hat{a}\right)}{\partial x}\right|_{x=\hat{x}} + \frac{\partial H\left(\hat{x},\hat{a}\right)}{\partial \hat{a}} \frac{\partial A\left(\hat{y},\hat{h}\right)}{\partial \hat{h}} \left.\frac{\partial h}{\partial x}\right|_{x=\hat{x}} && \Leftrightarrow \\ & \left.\frac{\partial h}{\partial x}\right|_{x=\hat{x}} = \frac{\left.\frac{\partial H\left(x,\hat{a}\right)}{\partial x}\right|_{x=\hat{x}}}{1-\beta_{\mathrm{H}}\,\beta_{\mathrm{B}}}, \end{aligned} \tag{S5}$$

where $\beta_{\mathrm{H}} = \frac{\partial}{\partial \hat{a}} H\left(\hat{x},\hat{a}\right)$ is the slope of the helper's reaction norm and $\beta_{\mathrm{B}} = \frac{\partial}{\partial \hat{h}} A\left(\hat{y},\hat{h}\right)$ is the slope of the breeder's reaction norm, evaluated at the equilibrium levels of help and aggression realized in interactions between residents.

Apart from quantifying the responsiveness of the helper and the breeder to variation in their partner's behavior, the slope parameters $\beta_{\mathrm{B}}$ and $\beta_{\mathrm{H}}$ also determine whether the negotiation process (S4) converges, and if so, how fast. An implicit assumption of derivation (S5) is that the negotiation between resident individuals attains a stable equilibrium $\left(\hat{a},\hat{h}\right)$. This is the case when the Jacobian matrix associated with the resident negotiation dynamics,

$$\mathbf{J} = \begin{pmatrix} 0 & \frac{\partial}{\partial \hat{a}} H\left(\hat{x},\hat{a}\right) \\ \frac{\partial}{\partial \hat{h}} A\left(\hat{y},\hat{h}\right) & 0 \end{pmatrix} = \begin{pmatrix} 0 & \beta_{\mathrm{H}} \\ \beta_{\mathrm{B}} & 0 \end{pmatrix}, \tag{S6}$$

has no eigenvalues with absolute values larger than 1. Accordingly, stability of the equilibrium $\left(\hat{a},\hat{h}\right)$ is ensured when $-1 < \beta_{\mathrm{B}}\,\beta_{\mathrm{H}} < 1$. Note that, under this same condition, the term in the denominator of the final result in (S5) is positive.

Similar to the calculation of $\frac{\partial h}{\partial x}$, it is possible to also obtain an explicit expression for $\frac{\partial a'}{\partial x}$, using the fact that $a' = A(\hat{y}, H(x,a'))$ at the negotiated equilibrium. The result of this derivation is that

$$\left.\frac{\partial a'}{\partial x}\right|_{x=\hat{x}} = \frac{\beta_{\mathrm{B}} \left.\frac{\partial H\left(x,\hat{a}\right)}{\partial x}\right|_{x=\hat{x}}}{1-\beta_{\mathrm{H}}\,\beta_{\mathrm{B}}}. \tag{S7}$$

Substituting results (S5) and (S7) into equation (S3) provides an expression for the selection gradient with respect to a genetic trait $x$ that affects the helper reaction norm:

$$w_x = \frac{\left.\frac{\partial H(x,\hat{a})}{\partial x}\right|_{x=\hat{x}}}{1-\beta_H\,\beta_B}\left(-c-\beta_{\mathrm{B}}\,d'\right). \tag{S8}$$

A similar analysis can be carried out for genetic traits that affect the breeder reaction norm, except that, in that case, the negotiation takes place between a helper expressing the resident reaction norm and a breeder expressing the mutant reaction norm. As a consequence, the relevant behavioral traits, $a$ and $h'$, are equilibria of the negotiation process

$$\begin{cases} h'_{t+1} = H\left(\hat{x}, a_t\right), \\ a_{t+1} = A\left(y, h'_t\right). \end{cases} \tag{S9}$$

Based on derivations similar to (S5) we eventually obtain the following result for the selection gradient with respect to genetic traits that affect the breeder's reaction norm

$$w_y = \frac{\left.\frac{\partial A\left(y,\hat{h}\right)}{\partial y}\right|_{y=\hat{y}}}{1-\beta_{\mathrm{H}}\,\beta_{\mathrm{B}}}\left(\beta_{\mathrm{H}}\,b-d\right). \tag{S10}$$

Several conclusions can be drawn from results (S8) and (S10). First, and as expected, there is no selection on genetic traits that do not change the behavioral phenotype of the helper ( $\frac{\partial}{\partial x}H\left(x,\hat{a}\right)\Big|_{x=\hat{x}} = 0$ ) or the breeder ( $\frac{\partial}{\partial y}A\left(y,\hat{h}\right)\Big|_{y=\hat{y}} = 0$ ). Second, since providing help is costly ($c > 0$) and receiving aggression causes costly damage as well ($d' > 0$), selection will only favor genes that increase helpfulness if resident breeders respond to increased levels of help by substantially reducing the amount of aggression they exert. A condition for the evolution of increased helping levels is that the slope of breeder reaction norm (which reflects the effective potential for appeasement) is sufficiently negative at the negotiated equilibrium, i.e.,

$$\begin{aligned} 0 &< -c-\beta_{\mathrm{B}}\,d' \\ \beta_B &< -\frac{c}{d'}. \end{aligned} \tag{S11}$$

This inequality is satisfied more easily if the marginal cost-benefit ratio of trading help for reduced aggression is favorable for helpers. The third general conclusion is that helpers must be sufficiently responsive to punishment in order for breeder aggression to evolve. In particular, genes that increase breeder aggression can spread only if the slope of the helper reaction norm satisfies

$$\beta_{\mathrm{H}} > \frac{d}{b}. \tag{S12}$$

The minimal level of responsiveness required to satisfy this condition is lower if the marginal damage-benefit ratio of inducing help by means of aggression is favorable for the breeder. On the other hand, since $d > 0$, the reaction norm of the breeder will evolve to minimize aggression if helpers are unresponsive.

*Evolution of help and aggression in a kin-structured population*

A helper who carries a mutation in the helper reaction norm, always interacts with a breeder expressing the resident reaction norm for aggression. However, in kin-structured populations, when the mutant individual becomes a breeder, it may interact with a helper that expresses the mutant reaction norm as well, due to the genetic correlations that exist between related helpers and breeders. In order to take into account these correlations, let us suppose that a mutant helper strategy $x$ interacts with a helper strategy $\tilde{x}$ at the time that the mutant becomes a breeder individual. The direction and the strength of selection on the mutant strategy is then measured by the gradient of the neighbor-modulated fitness[3,4] function with respect to $x$, given by

$$w_x = -c\left.\frac{\partial h}{\partial x}\right|_{x=\hat{x}} - d'\left.\frac{\partial a'}{\partial x}\right|_{x=\hat{x}} + \underbrace{\left.\frac{\partial \tilde{x}}{\partial x}\right|_{x=\tilde{x}=\hat{x}}}_{\substack{\text{relatedness}\\ \text{coefficient } r}}\left(b\left.\frac{\partial h'}{\partial \tilde{x}}\right|_{\tilde{x}=\hat{x}} - d\left.\frac{\partial a}{\partial \tilde{x}}\right|_{\tilde{x}=\hat{x}}\right). \tag{S13}$$

The four behavioral phenotypes that affect the fitness of the mutant are determined by the stable equilibrium of the negotiation dynamics

$$\begin{cases} \left.\begin{aligned} h_{t+1} &= H\left(x, a'_t\right) \\ a'_{t+1} &= A\left(\hat{y}, h_t\right) \end{aligned}\right\rangle \text{ interaction between mutant and related breeder ,} \\ \left.\begin{aligned} h'_{t+1} &= H\left(\tilde{x}, a_t\right) \\ a_{t+1} &= A\left(\hat{y}, h'_t\right) \end{aligned}\right\rangle \text{ interaction between mutant and related helper ,} \end{cases} \tag{S14}$$

which gives rise to expressions for the change of the behavioral traits in response to a change of $x$, similar to the derivation shown in (S5). Substituting these expressions into equation (S13) yields the result

$$w_x = \frac{\left.\frac{\partial H(x,\hat{a})}{\partial x}\right|_{x=\hat{x}}}{1-\beta_{\mathrm{H}}\,\beta_{\mathrm{B}}}\left(r\,b - c - \beta_B\left(d' + r\,d\right)\right). \tag{S15}$$

According to this result, if help has a positive marginal fitness effect at the level of the breeding group (i.e., $b - c > 0$), then help can evolve if the coefficient of relatedness

exceeds a critical marginal cost/benefit ratio $r > c / b$. Accordingly, in a kin-structured population, selection favors helping under a broader range of conditions, without necessarily requiring breeders to be responsive (cf. condition (S11) for a well-mixed population). From equation (S15) we also infer that increased help is favored by selection under the following condition on the slope of breeder reaction norm (this is inequality 2 in the main text):

$$\beta_{\mathrm{B}} < \frac{rb-c}{rd+d'}. \tag{S16}$$

The above derivation can be easily modified to calculate the strength of selection that acts on genes influencing the breeder-reaction norm. After the appropriate adjustments to equations (S9) and (S10), the result is:

$$w_y = \frac{\left.\frac{\partial A\left(y,\hat{h}\right)}{\partial y}\right|_{y=\hat{y}}}{1-\beta_H \beta_B}\left(\beta_H\left(b-rc\right)-d-rd'\right). \tag{S17}$$

In this case, increasing aggression is favored by selection only if

$$\beta_H > \frac{d+rd'}{b-rc}, \tag{S18}$$

which is inequality 1 in the main text. This condition is more difficult to satisfy than inequality (S12), because relatedness increases the breeder's marginal cost-benefit ratio of negotiation in two different ways: first, relatedness increases the inclusive fitness effect of the damage caused by aggression, because aggression by the breeder reduces the fitness of a relative; second, it also reduces the inclusive fitness benefit of help, because help received by the breeder comes at the cost of a fitness reduction for the related individual who is providing the help.

**Table 1. Summary of definitions used in the mathematical analysis**

| Symbol | Description |
|---|---|
| $a$ | Amount of aggression exerted by a mutant breeder |
| $a'$ | Amount of aggression exerted on a mutant helper |
| $\hat{a}$ | Breeder aggression exerted in interactions between residents |
| $h$ | Amount of help given by a mutant helper |
| $h'$ | Amount of help received by a mutant breeder |
| $\hat{h}$ | Amount of help provided in resident-resident interactions |
| $c=-w_h$ | Cost to the helper of providing help |
| $b=w_{h'}$ | Benefit to the breeder of receiving help |
| $d=-w_a$ | Cost to the breeder of being aggressive |
| $d'=-w_{a'}$ | Cost to the helper of receiving aggression |
| $\beta_{\text{sub}}=\frac{\partial}{\partial \hat{a}}H(\hat{x},\hat{a})$ | Slope of the helper reaction norm at the behavioral equilibrium |
| $\beta_{\text{dom}}=\frac{\partial}{\partial \hat{h}}A(\hat{y},\hat{h})$ | Slope of the breeder reaction norm at the behavioral equilibrium |
| $r\ (=\partial\tilde{x}/\partial x)$ | Coefficient of relatedness between breeder and helper |

## 2. Parameter choice and sensitivity analysis

We assessed the robustness of our main results, the existence of two alternative equilibria and the effect of philopatry on their frequencies, by changing most of the parameter values used in the simulations. Also, we made an alternative model where individuals reproduce sexually, and there is recombination of the two reaction norms.

For all the changes made in parameter values and type of reproduction, we recovered the existence of the "negotiated" and "kin selected" equilibria. Moreover, the frequency of the "negotiated" equilibrium decreased as philopatry ($\varphi$) was increased. The parameter values used for the simulations were inspired by the ecology and life-history of *Neolamprologus pulcher*. Field estimates of *N. pulcher's* yearly survival for dominant and subordinates are 0.56 and 0.44 respectively (Taborsky, *unpublish data*). In the model, average per time-step probability of survival for dominants at the end of the simulation is 0.59. As for helpers, average per time-step probability of survival changes depending of the evolutionary equilibria reached. In the negotiated equilibrium, helper survival is 0.48, while in the kin-selected equilibria is 0.57. Field estimates of yearly survival for floaters have not been reported. However, predation pressures are so high outside the territory that the value of 0.1 used in the model is not unrealistic. *N. pulcher* groups have an average of 5.5 helpers, and all of them together increase the productivity of a group by 70%. The average number of juveniles produced by one group in a year is 12.3 (Taborsky, *unpublish data*). In our model we assume, for the sake of simplicity, groups have only one helper. Therefore, the fecundity effect of a helper in the model corresponds to that of 5.5 helpers in *N. pulcher*. For consistency, we chose a baseline group productivity 5.5 times smaller than in field estimates. The evolved level of help in the model increased the fecundity of a dominant individual by 80%.

**Table** 2 shows the parameter values used in the simulations shown in the figures, as well as the range in which parameters were varied for the sensitivity analysis.

The parameter values used for the simulations were inspired by the ecology and life-history of *Neolamprologus pulcher*. Field estimates of *N. pulcher's* yearly survival for dominant and subordinates are 0.56 and 0.44 respectively (Taborsky, *unpublish data*). In the model, average per time-step probability of survival for dominants at the end of the simulation is 0.59. As for helpers, average per time-step probability of survival changes depending of the evolutionary equilibria reached. In the negotiated equilibrium, helper survival is 0.48, while in the kin-selected equilibria is 0.57. Field estimates of yearly survival for floaters have not been reported. However, predation pressures are so high outside the territory that the value of 0.1 used in the model is not unrealistic. *N. pulcher* groups have an average of 5.5 helpers, and all of them together increase the productivity of a group by 70%. The average number of juveniles

produced by one group in a year is 12.3 (Taborsky, *unpublish data*). In our model we assume, for the sake of simplicity, groups have only one helper. Therefore, the fecundity effect of a helper in the model corresponds to that of 5.5 helpers in *N. pulcher*. For consistency, we chose a baseline group productivity 5.5 times smaller than in field estimates. The evolved level of help in the model increased the fecundity of a dominant individual by 80%.

**Table 2. Parameters and evolutionary variables of the simulation model. The third column shows the parameter values used in the simulations presented in the figures. The last column shows ranges of alternative values used for the sensitivity analysis. Parameters that do not have alternative values, or a range, were kept constant in all simulations.**

| Symbol | Description | Simulations values | Sensitivity analysis |
|---|---|---|---|
| $N$ | Total number of territories | 10000 | 5000 |
| $S_0$ | Baseline survival probability on a territory | 0.6 | - |
| $S_F$ | Survival of floaters | 0.1 | 0.05, 0.5 |
| $k_0$ | Intercept of the logistic function determining dominant survival | -5 | - |
| $k_a$ | Marginal cost of aggression for the dominant | 1 | 0.1, 1 |
| $f_0$ | Intercept of the logistic function determining subordinate survival | -5 | - |
| $f_a$ | Marginal cost of aggression for the subordinate | 5 | 1, 5 |
| $f_h$ | Marginal cost of helping for the subordinate | 1 | 0.1, 1 |
| $F_0$ | Baseline fecundity | 2 | 2 - 5 |
| $B$ | Fecundity benefit of helping | 0.3 | 0.1 – 1.5 |

| | | | |
|---|---|---|---|
| $C$ | Fecundity cost of lazy subordinates | 0.1 | 0.1 - 0.25 |
| $\varphi$ | Probability that a local offspring becomes subordinate ("degree of philopatry") | 0 – 0.3 | 0 – 0.8 |
| $\eta$ | Competitive ability of subordinates | 2 | 1 - 15 |
| $\varepsilon_a$ | Environmental noise in the subordinate's decision to leave a territory | 0.01 | 0.01 – 0.3 |
| $\mu$ | Mutation rate | 0.005 | - |
| $\sigma_\mu$ | Width of the mutational step-size distribution | 0.2 | 0.1 – 0.3 |

Description of evolutionary variables

| | |
|---|---|
| $\tau$ | Tolerance threshold |
| $A_{\max}$ | Maximal aggression in the dominant reaction norm |
| $\alpha$ | Intercept of the dominants norm of reaction |
| $\beta$ | Weight of subordinate's helping level on dominant's behaviour |
| $H_{\max}$ | Maximal helping level in the subordinate reaction norm |
| $\gamma$ | Intercept of the subordinate reaction norm |
| $\delta$ | Weight of dominant's aggression level on subordinate's behaviour |

**Cited literature**

1. Pen, I. & Weissing, F. J. Towards a unified theory of cooperative breeding: the role of ecology and life history re-examined. *Proc. R. Soc. B Biol. Sci.* **267,** 2411–2418 (2000).
2. Caswell, H. *Matrix Population Models*. (Sinauer Associates, 2000).
3. Taylor, P. D., Wild, G. & Gardner, A. Direct fitness or inclusive fitness: how shall we model kin selection? *J. Evol. Biol.* **20,** 301–309 (2007).
4. Taylor, P. D. & Frank, S. A. How to Make a Kin Selection Model. *J. Theor. Biol.* **180,** 27–37 (1996).

Evolutionary invasion analysis - Mathematical derivation of the expressions discussed in the text

Figure S1 - Transition from a kin-selected equilibrium to a negotiated equilibrium. (A) Evolutionary dynamics of aggression (red line, left axis) and helping level (blue line, right axis) in a simulation run with a high degree of relatedness between breeders and helpers (r=0.26). (B) Reaction norms of breeders (red) and helpers (blue) at five different points in time. (B1) The simulation starts with unresponsive strategies that provide no help, impose no aggression, and are not tolerant to aggression. (B2) The population has reached a kin-selected equilibrium where breeders are unresponsive at the intersection point of the two reaction norms. helpers provide help without the need of aggression being enforced. The reaction norms of the helpers show considerable variation in their responses to higher aggression levels, since these aggression levels do not occur and therefore do not exert a selection pressure. (B3) At some point in the simulation, the average reaction norm of the helpers (thick blue line) becomes an increasing function, allowing the breeders to receive more help by increasing the level of aggression. After both breeders and helpers have become responsive to each other (B4), the population rapidly converges to the negotiated equilibrium (B5). At the new equilibrium, the helping level of helpers is considerably higher than at the kin-selected equilibrium, although aggression levels remain very low.

Figure S2 - Effect of an increase of philopatry (and relatedness) on the evolved levels of helping and aggression in negotiated equilibria (purple) and kin-selected equilibria (green). For four degrees of philopatry ( ), which generated different levels of relatedness ( ), the outcome of 50 replicate simulations was classified as either ‘negotiated’ or ‘kin-selected’, see caption in Fig. 5. Dots indicate the average helping and aggression level per category; error bars represent standard errors. Numbers below the data points indicate the number of simulation runs that were classified in ‘negotiated’ and ‘kin-selected’ equilibria.

Figure S3 - Effect of an increase in the phenotypic effect size of mutations on the evolved levels of helping and aggression in negotiated equilibria (purple) and kin-selected equilibria (green). For three different standard

deviations of the normal distribution from which mutational effect sizes are drawn ($s_\mu$={0.1,0.2,0.4}), the outcome of 20 replicate simulations was classified as 'negotiated' or 'kin-selected'. For each simulation, the levels of help and aggression were averaged over the last 500 generations; error bars represent standard errors. (A) when the mutational effect-size distribution is more narrow (i.e., mutations have smaller phenotypic effects), relatively many simulations are trapped in the 'kin-selected' equilibrium. As the distribution becomes wider (B and C), more simulations escape the 'kin-selected' equilibria and reach the more productive negotiated equilibrium. This result suggests that mutations with large phenotypic effect are needed to jump from the 'kin-selected' equilibrium into the basin of attraction of the 'negotiated' equilibrium.

Figure S4 - Sensitivity analysis of model predictions for three parameter values: benefits of help (B), cost of a subordinate (C) and survival of floaters (SF). Each point represents the evolved level of help and aggression in one replicate simulation. For each simulation, the levels of help and aggression were averaged over the last 500 generations; error bars represent standard errors. Levels of help and aggression evolved under the absence of philopatry (red dots) and high philopatry (green dots $\varphi$=1/3). Changes in the three parameters influence the location of the two equilibria and the relative frequencies of simulations that end up in either of them, but the existence of the two equilibria is robust to changes in the parameter values.

A

Aggression

Helping level

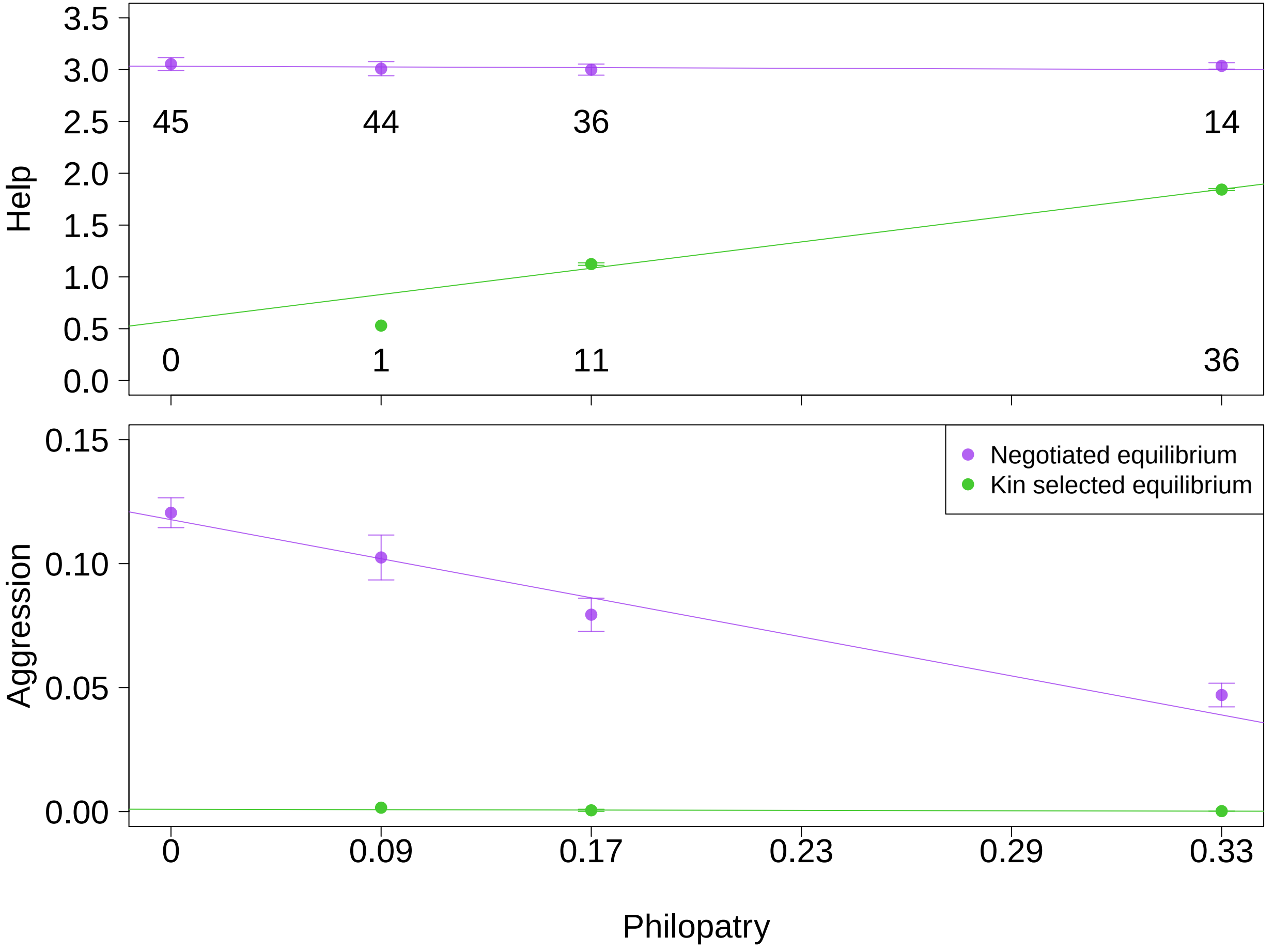
Generations

1 2 3 4 5

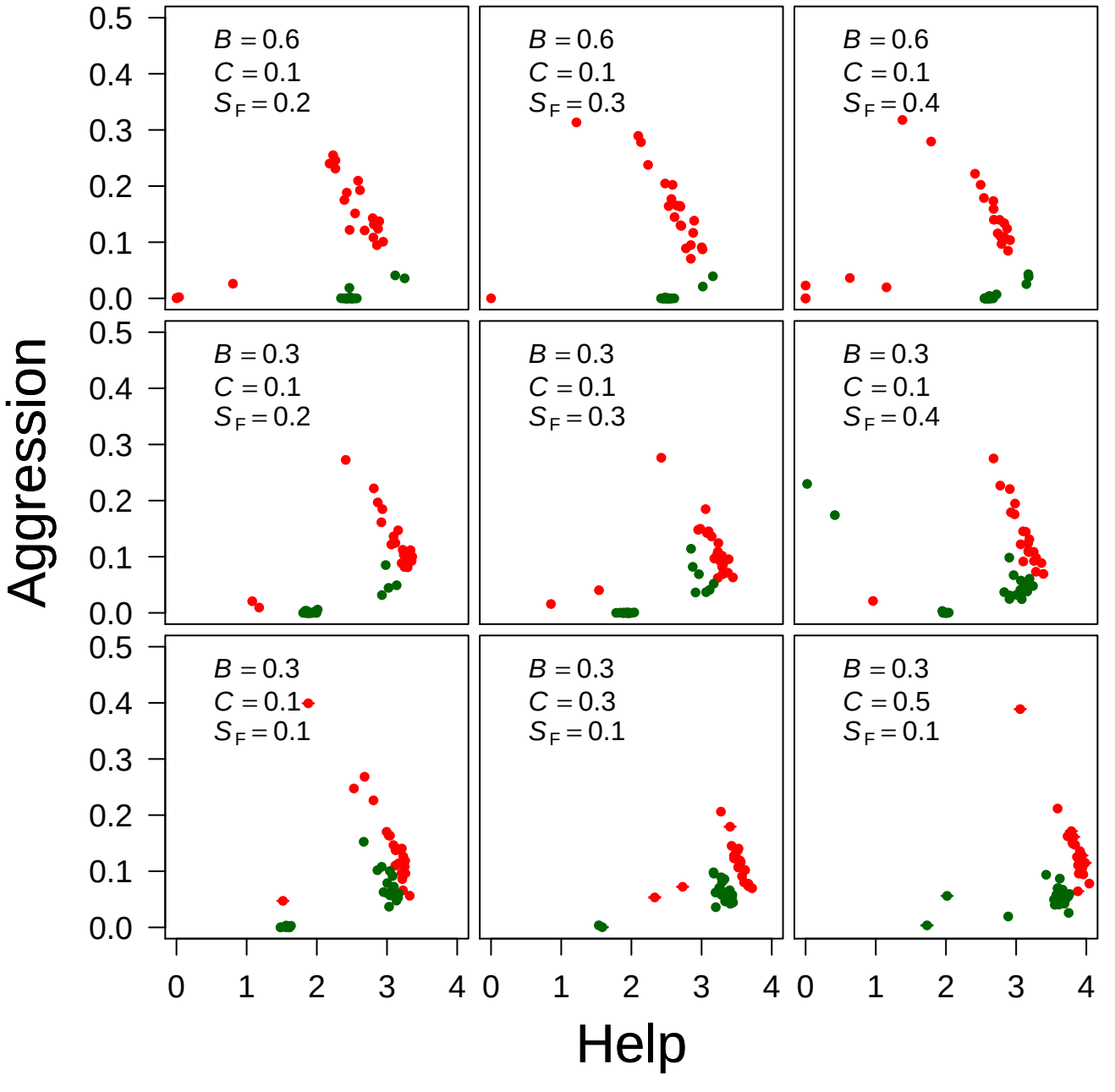
B

Breeder
Helper

Helping level


Help
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
45
44
36
14
0
1
11
36
Aggression
0.15
0.10
0.05
0.00
0
0.09
0.17
0.23
0.29
0.33
Philopatry
Negotiated equilibrium
Kin selected equilibrium

A
$\sigma_\mu = 0.1$
B
$\sigma_\mu = 0.2$
C
$\sigma_\mu = 0.4$
Aggression
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
6
11
20
14
9
0
1
2
3
4
Helping behaviour


Aggression
Help
B = 0.6
C = 0.1
S_F = 0.2
B = 0.6
C = 0.1
S_F = 0.3
B = 0.6
C = 0.1
S_F = 0.4
B = 0.3
C = 0.1
S_F = 0.2
B = 0.3
C = 0.1
S_F = 0.3
B = 0.3
C = 0.1
S_F = 0.4
B = 0.3
C = 0.1
S_F = 0.1
B = 0.3
C = 0.3
S_F = 0.1
B = 0.3
C = 0.5
S_F = 0.1
0.0
0.1
0.2
0.3
0.4
0.5
0
1
2
3
4