数理解析研究所講究録
第 827 巻 1993 年 165-175



# AN EVOLUTIONARY THEORY OF CONFLICT RESOLUTION BETWEEN RELATIVES: ALTRUISM, MANIPULATION, COMPROMISE

Norio Yamamura (山村則男)
Department of Natural Sciences, Saga Medical School,
Nabeshima, Saga 849, Japan
and
Masahiko Higashi (東正彦)
Faculty of Science and Technology, Ryukoku University,
Seta, Otsu 520-21, Japan

Many cases of conflict between relatives over the evolution of social behavior are known (e. g., Dawkins, 1976; Trivers, 1985), but no general theory incorporates the consequences of conflict. Here we show how compromise solutions may evolve. We derive the compromise solution by incorporating conflict costs into the fitness evaluation. Specifically, for donor-recipient conflict concerning altruism, we find that altruism evolves more easily than Hamilton's Rule (Hamilton, 1964) would predict for the case with no effective manipulation by the recipient, whereas in the opposite case, where the donor cannot resist the recipient's manipulation, the behavior evolves less easily than the Inverse Hamilton's Rule would predict. The theory also indicates conditions under which no compromise is reached, and physical conflict is manifested.

For the Hymenoptera, it has been debated whether the evolution of sterile worker castes is due to kin selection for daughters' altruistic behavior toward their mother (Hamilton, 1964), or to the development of the mother's manipulation which forces "unwilling" daughters to serve her (Alexander, 1974; Charnov, 1978). Trivers (1974) considered the problem of parent-offspring conflict over parental investment. In many cases besides these the existence of conflict has been stressed, but there has been remarkably little work on the important problem of where the conflict should lead. As in the case of mother-daughter conflict in the evolution of sterile workers, settlement of the conflict has been considered to be simply a matter of conquest by the stronger (e.g., mother's manipulation). But both sides engaged in any conflict are expected to pay some cost, and if the costs are taken into account, it seems that the outcome should be affected by not only the relative strengths of the players, but also by their relative benefits (or losses) when they win (or lose) the conflict.

Here we present a general scheme by which rules may be derived for the resolution of conflicts, thereby identifying factors that determine the outcome of the conflict. A key idea in this scheme is the incorporation of conflict costs into the inclusive fitness evaluation. We illustrate the general scheme, using as a model case the donor-recipient conflict over the evolution of altruistic behavior.

Consider the evolution of an altruistic behavior by the donor (hereafter referred to as **D**) that decreases its own fitness by $C$ and increases by $B$ the fitness of the recipient (hereafter **R**). A symmetric genetic relation between **D** and **R** is assumed with degree of relatedness $r$ . The condition for the altruistic behavior of **D** (without **R**'s control) to be favoured by selection is that the inclusive fitness of **D** is greater when it performs this altruistic behavior than when it does not: $W_{\mathbf{D}} - C + rB > W_{\mathbf{D}}$, where $W_{\mathbf{D}}$ is **D**'s fitness without any social interaction (Hamilton, 1964). From this, Hamilton's Rule immediately follows:

$$\frac{B}{C} > \frac{1}{r} \quad . \qquad (1)$$

On the other hand, the condition for this altruistic behavior to increase **R**'s inclusive fitness is $W_{\mathbf{R}} + B + r(-C) > W_{\mathbf{R}}$, where $W_{\mathbf{R}}$ is **R**'s fitness without any social interaction. From this follows a different relationship,

$$\frac{B}{C} > r \quad , \qquad (2)$$

which may be called the *Inverse Hamilton's Rule*. Even if the recipient could force the donor to perform the altruistic behavior, the recipient should not do so when condition (2) does not hold.

The gap between conditions (1) and (2) implies a conflict. If $B/C$ lies between $r$ and $1/r$ , *i.e.*,

$$r < \frac{B}{C} < \frac{1}{r} \, , \qquad (3)$$

then **D** should not perform the altruistic behavior [because condition (1) does not hold], whereas **R** should attempt to make **D** perform the altruistic behavior [because condition (2) does hold] (see Fig. 1). Notice that the conflict region is reduced with $r$, banishing when $r = 1$; there exists no conflict between genetically identical individuals. On the contrary, when $r = 0$, the conflict region expands to the whole region of positive values of $B$ and $C$.

However, if costs involved in the conflict are taken into account in the inclusive fitness evaluation, then this potential conflict may have an evolutionary resolution. Here, by costs involved in the conflict, we mean reduction in the fitness of the recipient in manipulating the donor and that of the donor in resisting against the recipient's control. There exists in the $(C, B)$-space a critical line that divides the conflict region defined by condition (3) into two sub-regions. In the sub-region above the line the donor is selected to avoid conflict by performing the altruistic behavior, whereas in the sub-region below the line the recipient should avoid conflict by not attempting to manipulate the donor. This compromise line for the donor-recipient conflict is

$$\frac{B}{C} = \frac{2kr + r^2 + 1}{k(r^2 + 1) + 2r} \quad . \tag{4}$$

A derivation follows.

Suppose that parameters $(C, B)$ take values that fall in the conflict region defined by condition (3). Let $d_{\mathbf{D}}$ and $d_{\mathbf{R}}$, respectively, denote the costs paid by **D** and **R** in pursuing the conflict, and assume that **R** will dominate **D** in the conflict when $kd_{\mathbf{R}} > d_{\mathbf{D}}$, whereas **D** will dominate **R** when $kd_{\mathbf{R}} < d_{\mathbf{D}}$. That is, in order for **D** to resist **R**, it has to pay $k$ times as much cost as **R** does; thus, $k$ represents the degree of dominance of **R** over **D** in fighting the conflict using the same amount of cost.

Then, **D** and **R** will evolve to increase their conflict costs in order to rival each other. When they build their costs up to $d_{\mathbf{D}}$ ($=kd_{\mathbf{R}}$) and $d_{\mathbf{R}}$, respectively, the inclusive fitness that **D** would earn in the case of winning the conflict should be reduced to $W_{\mathbf{D}} - kd_{\mathbf{R}} + r(-d_{\mathbf{R}})$, and that of **R** to $W_{\mathbf{R}} - d_{\mathbf{R}} + B + r(-C - kd_{\mathbf{R}})$. (Note that the cost of conflict paid by one side contributes a negative effect to not only its own fitness, but also to the other's inclusive fitness through their relatedness $r$.) As the conflict costs ($d_{\mathbf{R}}$) increase, the inclusive fitness of **D** (or **R**) will decline toward the value that it would take if it yielded to the other in the first place. Eventually, the better choice for **D** (or **R**) is to yield to **R** (or **D**), and the altruistic behavior will (or will not) evolve.

The condition that **D**'s inclusive fitness value in the case of winning is greater than it would be in the case of avoiding the conflict in the first place is given as $W_{\mathbf{D}} - kd_{\mathbf{R}} + r(-d_{\mathbf{R}}) > W_{\mathbf{D}} - C + rB$, or equivalently

$$\frac{B + d_{\mathbf{R}}}{C - kd_{\mathbf{R}}} < \frac{1}{r} . \qquad (5)$$

The same condition for **R** is given as $W_{\mathbf{R}} - d_{\mathbf{R}} + B + r(-C - kd_{\mathbf{R}}) > W_{\mathbf{R}}$, or equivalently

$$\frac{B - d_{\mathbf{R}}}{C + kd_{\mathbf{R}}} < r . \qquad (6)$$

[For a geometric interpretation in $(C, B)$-space of conditions (5) and (6), see Fig. 1 and its legend.]

Therefore, the condition that **D** should yield to **R** and perform the altruistic behavior is that for some $d_{\mathbf{R}}$, both sides of condition (5) become equal to each other, while inequality (6) still holds. Eliminating $d_{\mathbf{R}}$ from these two conditions, we get

$$\frac{B}{C} > \frac{2kr + r^2 + 1}{k\,(r^2 + 1) + 2r} \equiv g(r, k)\,. \qquad (7)$$

Likewise, the condition for **R** to yield to **D** and not attempt to induce the altruistic behavior is given by inequality (7) with the opposite inequality sign (<). Therefore, the compromise line, which represents the critical condition for the evolution of the altruistic behavior, is given by equation (4) (Fig. 2).

The slope of the compromise line, $g(r, k)$, decreases with increasing $k$ ($0 < k < \infty$) and is confined between $g(r, 0) = (r^2 + 1)/2r$ and $g(r, \infty) = 2r/\,(r^2 + 1)$:

$$r \leq \frac{2r}{r^2+1} \leq g(r,k) \leq \frac{r^2+1}{2r} \leq \frac{1}{r} \qquad (8)$$

where $0 \leq r \leq 1$ and the equalities of both ends hold only when $r = 1$. The highest and lowest values of $g(r, k)$ are the arithmetic mean and the harmonic mean, respectively, of $r$ and $1/r$. Thus, even for the extreme case ($k \rightarrow 0$ or $\infty$) the compromise line does not reach the boundary lines of the conflict region unless $r = 1$, as seen in Fig. 2. The reason for this is as follows. Even if **R** has to pay virtually no cost ($d_R = d_D/k \rightarrow 0$) due to its absolute dominance in control ($k \rightarrow \infty$), the cost $d_D$ paid by **D** (for the resistance against **R**) puts a burden on **R**'s inclusive fitness, $W_R + B + r\,(-C - d_D)$, through their relatedness ($r$). Similarly, even if **D** has to pay virtually no cost ($d_D = kd_R \rightarrow 0$) due to its absolute dominance in resistance ($k \rightarrow 0$), the cost $d_R$ paid by **R** (in the efforts of retaining **D** for the altruistic service toward **R**) puts a burden on **D**'s inclusive fitness, $W_D + r(-d_R)$, through their relatedness ($r$). This fact has an implication that demands a revision of Hamilton's Rule and the Inverse Hamilton's Rule, unless conflict costs are all negligible, as we see in the following.

First, for the case of no effective manipulation by **R** ($k \to 0$), the threshold value of $B/C$ for the altruistic behavior to evolve for each value of $r$,

$$\frac{B}{C} = g(r,\ 0) = \frac{r^2+1}{2r}, \qquad (9)$$

is much lower than the one that Hamilton's Rule predicts (Fig. 3). For example, when $r = 1/2$ and $1/4$, respectively, Hamilton's Rule predicts that $B/C$ must be 2.0 and 4.0 for the altruistic behavior to evolve, whereas the new theory predicts that it must be more than only 1.25 and 2.125. Or when $B/C$ is 2.0, Hamilton's Rule predicts that $r$ must be more than 1/2 for the altruistic behavior to evolve, whereas the new theory predicts that it must be more than only 0.268.

For the case of no effective resistance by **R** ($k \to \infty$), the threshold value of $B/C$ for the altruistic behavior to evolve is

$$\frac{B}{C} = g(r,\ \infty) = \frac{2r}{r^2+1} \quad . \qquad (10)$$

This is much higher than the $B/C$ ratio predicted by the Inverse Hamilton's Rule predicts (Fig. 3). For example, when $r = 1/4$ and $1/2$, respectively, the Inverse Hamilton's Rule predicts that $B/C$ must be only 0.25 and 0.5 for the altruistic behavior to evolve, whereas the new theory predicts that it must be more than 0.471 and 0.8. Or when $B/C$ is 0.5, the Inverse Hamilton's Rule predicts that $r$ must be less than 1/2 for the altruistic behavior to evolve, whereas the new theory predicts that it must be less than 0.268.

Finally, the theory developed here indicates under what conditions no compromise is reached, and physical conflict may ensue. In general, the parameters $B$, $C$, and $k$ are not constant in specific conflict scenes. In such a case, both players **D** and **R** will evolve to adopt conditional strategies, depending on the present values of $B$, $C$, and $k$. If assessment of the parameter values by both players is exactly correct, a compromise will be reached, as derived above. If one of the two incorrectly estimates the parameters and if both think they can win the conflict, then the conflict will actually start. Specifically, when the parameter values are close to the compromise line, then the judgment, whether to fight or not, will be difficult. Therefore, we can say that the closer the parameter values are to the compromise line (*i.e.,* satisfying the critical condition (4) for the evolution of altruism), the more likely it is that no compromise will be reached, and physical conflict will be actually manifested.

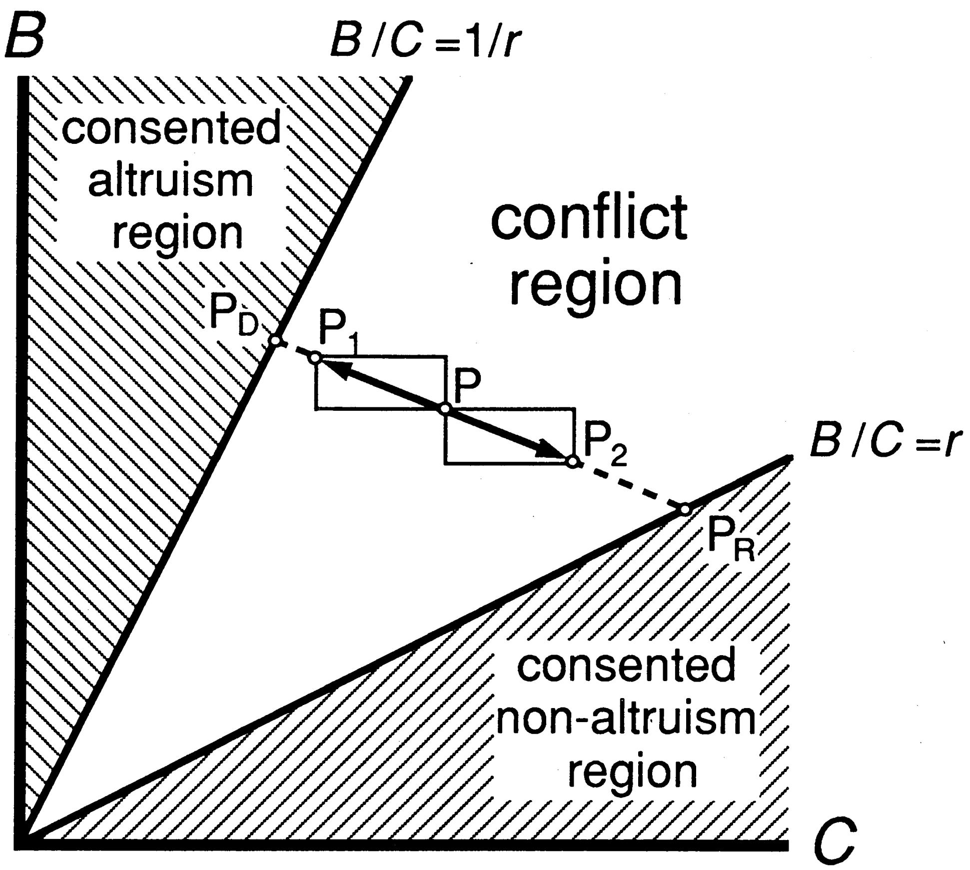


Fig. 1 Three regions in (C, B)-space with different implications for the evolution of an altruistic behavior of donor (D) toward recipient (R). The middle region defined by condition (3) is the region of conflict between D and R. The outer regions defined by inequalities B/C > 1/r and B/C < r are the regions of consensus between D and R : the former is the altruism region, where in the sense of evolutionary choice, R "wants D", and D "is willing", to take the altruistic behavior, while the latter is the non-altruism region, where D "wants", and R "is willing to permit D", not to take the altruistic behavior. When point P(C, B) falls in the conflict region, D and R both would have to pay costs if they pursued the conflict. The condition that D's inclusive fitness value in the case of winning is greater than it would be in the case of avoiding the conflict in the first place, given as inequality (4), can be interpreted as being that P1(C-kdR, B+dR), the point obtained by shifting P(C, B) with vector (-kdR, dR), is still located below line B/C = 1/r . The same condition for R, given as inequality (5), is that P2(C+kdR, B-dR), the point obtained by shifting P(C, B) with vector (kdR, -dR), is still located above line B/C = r. As dR (the conflict cost) increases, the varying points P1 and P2 approaches the critical lines B/C = 1/r and B/C = r, respectively. Let PD and PR, respectively, denote the intersection points that P1 and P2 meet when they reach those lines. Then, the condition for D to lose the conflict game can be stated that P1 reaches PD, while P2 does not yet reach PR, that is, P is closer to PD than it is to PR.

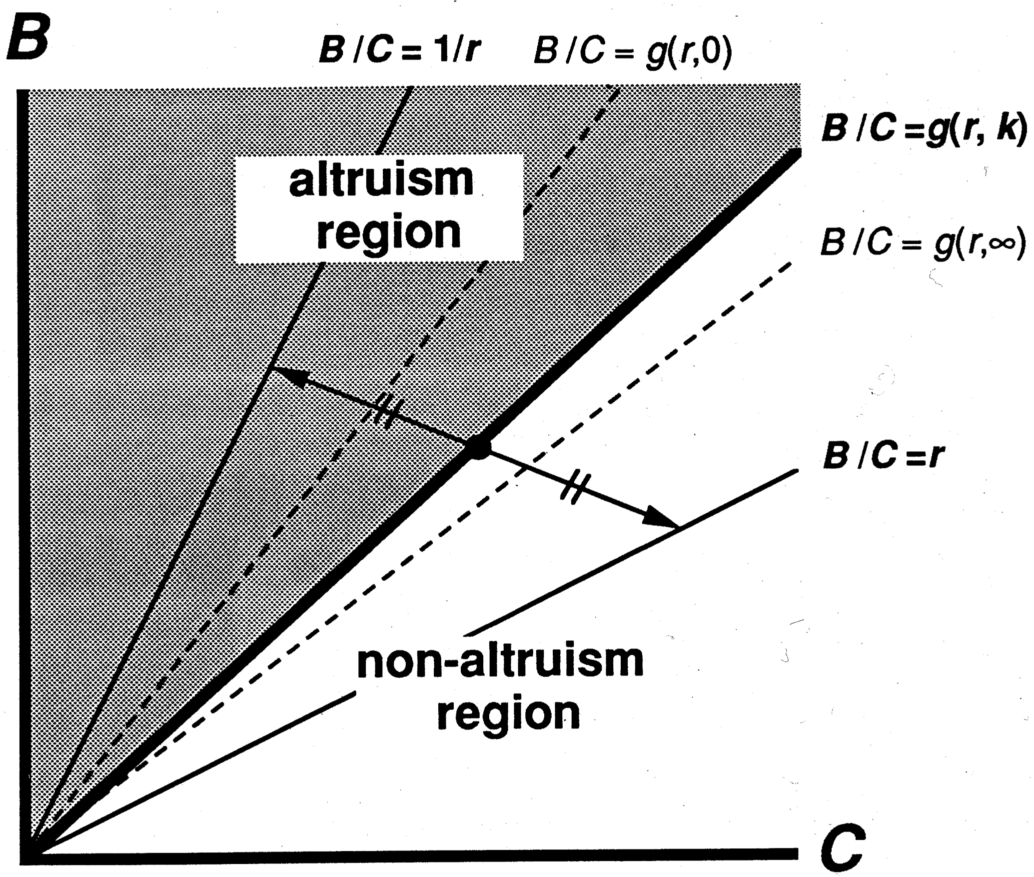


Fig. 2 Summary of the results for conflict resolution in the (C, B)-space. Line B/C = g(r, k) divides the conflict region into two sub-regions: the upper is the compromised altruism region, where D should withdraw from the conflict and take the altruistic behavior; and the lower is the compromised non-altruism region, where R, on the contrary, should withdraw from the conflict and let D not take the altruistic behavior. Notice that this line consists of the midpoints of the line segments (with slope -1/k) defined by the pairs of intersection points (PD and PR) indicated in Fig. 1. The slope of the compromise line, g(r, k), decreases as k increases (0 ≤ k ≤ •), and its range is confined between two broken lines (symmetric to each other with respect to line B/C = g(r, 1) = 1) whose slopes are g(r, 0) = (r2 + 1)/2r and g(r, •) = 2r/ (r2 + 1), respectively.

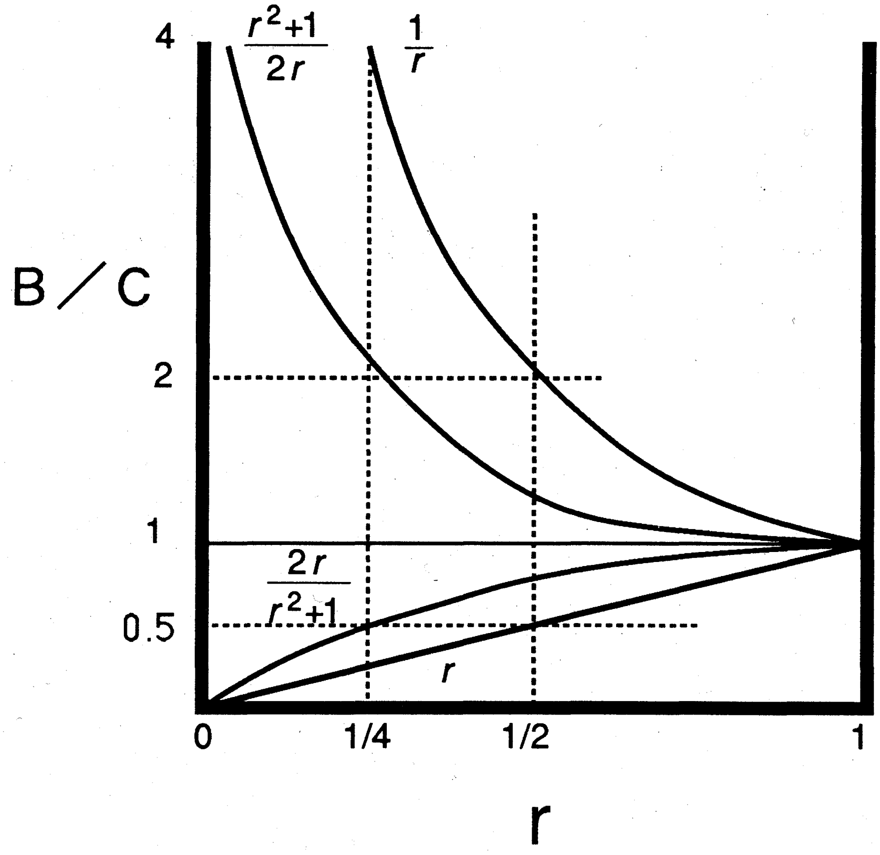


Fig.3 The lines representing the critical conditions for the altruism evolution represented as the relations of B/C to r that Hamilton's Rule predicts (the first line from the top), the new theory predicts for the case of no effective manipulation by R (k Æ 0) (the second line from the top), it predicts for the case of no effective resistance by R (k Æ •) (the second line from the bottom), and Inverse Hamilton's Rule predicts (the first line from the bottom).