Conflicts of interest and the evolution of decision sharing

Larissa Conradt* and Timothy J. Roper

Department of Biology & Environmental Sciences, University of Sussex, JMS Building, Falmer, Brighton BN1 9QR, UK

Social animals regularly face consensus decisions whereby they choose, collectively, between mutually exclusive actions. Such decisions often involve conflicts of interest between group members with respect to preferred action. Conflicts could, in principle, be resolved, either by sharing decisions between members (‘shared decisions’) or by one ‘dominant’ member making decisions on behalf of the whole group (‘unshared decisions’). Both, shared and unshared decisions, have been observed. However, it is unclear as to what favours the evolution of either decision type. Here, after a brief literature review, we present a novel method, involving a combination of self-organizing system and game theory modelling, of investigating the evolution of shared and unshared decisions. We apply the method to decisions on movement direction. We find that both, shared and unshared, decisions can evolve without individuals having a global overview of the group’s behaviour or any knowledge about other members’ preferences or intentions. Selection favours unshared over shared decisions when conflicts are high relative to grouping benefits, and vice versa. These results differ from those of group decision models relating to activity timings. We attribute this to fundamental differences between collective decisions about modalities that are disjunct (here, space) or continuous (here, time) with respect to costs/benefits.

Keywords: collective decisions; conflict resolution; consensus decisions; cooperation; information pooling; speed versus accuracy trade-off

1. GENERAL INTRODUCTION

A major challenge to most social animals, including humans, is the regular need to make ‘consensus decisions’, i.e. to choose, collectively, between mutually exclusive actions (Conradt 1998). Such decisions may involve choosing between different activities (Conradt 1998; Ruckstuhl & Neuhaus 2000), movement destinations (Trillmich et al. 2004; Couzin et al. 2005), nest sites (Seeley & Buhrman 1999), migration routes (Simons 2004) or hunting strategies (Conradt & Roper 2005).

In principle, three main factors influence consensus decisions, namely, information, conflict and time constraints. Firstly, the relevant information held by individual group members, and how this information is shared within the group, influences decision outcomes and accuracy (Seeley & Buhrman 1999; Conradt & Roper 2003; List 2004; Simons 2004; Couzin et al. 2005; Ame et al. 2006; Biro et al. 2006; Passino & Seeley 2006; Codling et al. 2007; Lusseau 2007; Ward et al. 2008; Dyer et al. 2009; Franks et al. 2009; List et al. 2009; Skyrms 2009; Sumpter & Pratt 2009). Secondly, conflicts of interest between group members, in the form of conflicting preferences for different decision outcomes, affect the fitness of individual group members and the evolution of the decision-making process (Conradt 1998; Ruckstuhl 1998, 1999; Conradt & Roper 2000, 2003, 2005, 2007; Ruckstuhl & Neuhaus 2000, 2002; Rands et al. 2003; Austen-Smith & Feddersen 2009; Hix & Noury 2009). Thirdly, time constraints and the resultant trade-offs (e.g. between speed and accuracy of decision making) limit the options for potential decision outcomes (Passino & Seeley 2006; Franks et al. 2009; Sumpter & Pratt 2009). Here, we concentrate on the second of these three factors, namely, conflicts of interest between group members.

To reach a consensus, conflicts of interest with respect to decision outcomes have to be resolved (Conradt & Roper 2003, 2007; Rands et al. 2003; Couzin et al. 2005). In principle, this can be achieved in two ways: decisions can either be ‘shared’ (some or all group members contribute to the decision outcome) or ‘unshared’ (the decision outcome is dictated by a single individual; Conradt & Roper 2003; List 2004; Austen-Smith & Feddersen 2009). Shared and unshared decisions influence the fitness of individual group members differently. Here, we are interested in the conditions under which shared decision making might be expected to evolve when there are conflicts of interest between group members. We provide a brief overview of published work on conflicts of interest and the evolution of decision sharing in animals and then investigate the evolution of decision sharing using a novel approach in which we integrate self-organizing systems simulations with game theory modelling.

2. SHORT REVIEW: CONFLICTS OF INTEREST AND THE EVOLUTION OF DECISION SHARING

Consensus decisions in animals vary in the extent to which they involve conflicts of interest between group members (Conradt & Roper 2005). For example, nest
site choice in eusocial insects (Seeley & Buhrman 1999; Visscher & Seeley 2007; Franks et al. 2009), or navigation in homing or migrating birds (Wallraff 1978; Simons 2004; Biro et al. 2006), involve relatively little conflict of interest because the optimal decision outcome is the same for all group members. By contrast, synchronization of group activities and choice of short-term travel destinations in birds, fishes and mammals can involve significant conflict (Conradt 1998; Ruckstuhl & Neuhaus 2000), since individual group members often differ in physiological requirements (e.g. Clutton-Brock et al. 1982) and therefore in their optimal requirements with respect to activity timing and travel destinations (e.g. Gompper 1996; Prins 1996; Ruckstuhl 1998).

Where there is conflict, reaching a consensus decision will necessarily involve compromise for at least some group members. Formally, this means that at least some members will pay a ‘consensus cost’, defined as the fitness cost to an individual of foregoing its own optimal action in order to comply with the consensus decision (Conradt & Roper 2003, 2005, 2007). Consensus costs, and their distribution between group members, are crucial for group stability (Conradt 1998; Ruckstuhl & Neuhaus 2000).

The consensus cost to individual group members depends on their own ability to influence the decision outcome versus that of other group members (Conradt & Roper 2003). For example, if a dominant individual dictates the decision (unshared decision making), it will not incur a cost because it can choose the decision outcome that is optimal for itself, whereas subordinate members will incur costs according to how different their own optima are from that of the dominant. By contrast, if all group members contribute equally to the decision outcome (‘equally shared’ decision making) the consensus costs will be relatively lower for most members other than the dominant (Conradt & Roper 2003). Decision-making processes that are intermediate between these two extremes (i.e. ‘partially shared’ decisions in which a subset of group members contributes to the decision outcome) are also possible.

Historically, scientists working on relatively small groups (i.e. groups in which global communication between group members is possible; Conradt & Roper 2005), especially in mammals, have tended to assume that decision making is unshared, no doubt because such groups usually exhibit obvious dominance hierarchies (e.g. Sapolsky 2005). By contrast, scientists working on large groups (i.e. groups in which only local communication between neighbouring members is possible; Conradt & Roper 2005), especially those using self-organizing system approaches, have tended to assume equally shared decision making insofar as they apply similar behavioural parameters to all group members (see Couzin & Krause 2003 for a review). The reason for this is no doubt the intrinsic difficulty of imagining that any individual member of a large self-organizing group could exert meaningful control over the behaviour of the group as a whole.

However, emerging empirical evidence challenges the assumption that a dominant individual always dictates what happens in small groups of animals (primates: Stolba 1979; Stewart & Harcourt 1994; Boinski & Campbell 1995; Byrne 2000; Milton 2000; Meunier et al. 2006; Sellers et al. 2007; ungulates: Conradt & Roper 2003; bats: Kerth et al. 2006; birds: Black 1988). Conversely, there is no intrinsic reason why self-organizing systems should only produce equal sharing of decisions (Conradt et al. in press). Indeed, several empirical observations suggest that in self-organizing groups individuals that have greater ‘needs’ might be able to gain a disproportionate weight in decision outcomes (fishes: Krause et al. 1992; Krause 1993; ungulates: Prins 1996; Fischhoff et al. 2007).

One of the most important outstanding theoretical questions is to explain how shared decision making can evolve in decisions that involve conflict, without invoking ‘good-for-the-group’ arguments (Gächter & Herrmann 2009). The main challenge in addressing this question is the complexity arising from multi-individual interactions in group decisions. Rands et al. (2003) modelled a group of two in which each member tries to maximize its own chance of survival and both gain from foraging together, but in which the optimal times for foraging can differ between individuals. In such a situation, if one member starts to forage because its reserves drop below a crucial threshold, the other member should also start to forage, because the advantages of foraging together outweigh the disadvantages of foraging at a less preferred time. The result is high activity synchrony. This constitutes, in principle, a shared consensus decision, because neither of the members has intrinsically more weight in the decision than the other, or differs in principle from the other at the start of the simulations. The decision threshold in this shared decision is one since, if at least one of the two starts to forage, this determines the collective behaviour. However, because group members differ in reserves, and this difference is likely to be preserved for some time, it is likely that during several consecutive occasions it will be the same individual that starts to forage first. Thus, this individual will seem to be the apparent ‘leader’ of the activity. It is not clear what would happen if one were to extend Rands et al.’s model to a group of more than two.

Dostalkova & Spinka (2007) also modelled activity synchronization between two group members. Similar to Rands et al. (2003), they predict that individuals maximize their benefits by switching activities simultaneously when the first prefers a change, provided that there is no communication between members. However, if there is communication, and acting too late is ‘cheaper’ (in terms of consensus cost) than acting prematurely, the group should switch activity only when both members prefer to do so (i.e. when a decision threshold of two is reached). In order to explain this difference, Dostalkova & Spinka (2007) raised the interesting notion of the inherent informational asymmetry of time. Without communication, information about members’ preferred leaving times becomes available only once members attempt to leave, resulting in an informational bias towards earlier preferred leaving times. This can affect synchronization decisions in an asymmetric manner. Again, it is not clear what will happen if the model is extended to a group larger than two.
A model by Conradt & Roper (2003), again focusing on activity synchronization, suggests that the consensus costs to a group as a whole are considerably lower for equally shared than for unshared decisions under most conditions and for group sizes larger than two. However, without invoking group selection arguments this does not necessarily lead to the evolution of decision sharing. Indeed, Conradt & Roper (2007) showed that, in principle, unshared decisions, as well as equally shared decisions, can be evolutionarily stable in groups of all sizes. The authors identified five factors that favour the evolution of shared versus unshared decisions in the context of activity synchronization, namely, (i) large conflicts of interest between group members (resulting in high potential consensus costs), (ii) long-term differences between group members with respect to requirements (i.e. high heterogeneity within the group), (iii) asymmetric consensus costs (i.e. one decision outcome is potentially much more costly than another), (iv) marginal grouping benefits, and (v) a group size that is close to, or above, optimal group size.

The predictions of all three of these models still require empirical testing. Moreover, all three models deal with activity synchronization, leaving it unclear as to whether the same results apply to the evolution of consensus decisions involving, say, movement destinations. Decisions about the latter often differ in principle from those about activity synchronization in an important respect, as follows. While an average or median timing might often offer a sensible compromise between different optimal timings (Conradt & Roper 2003), such a procedure would more rarely be valid when compromising between different spatial targets (Conradt et al. in press). For example, if a flock of geese were trying to decide which of two lakes to land on, it would make no sense for it to land halfway between the two lakes where there was no water at all. Another limitation of all three models is that they assume (at least implicitly) that group members have a global overview of the activities of other group members, and that this global overview can influence their behaviour. This assumption does not hold in many of the large groups that we observe in nature (Couzin & Krause 2003; Couzin et al. 2005; Sumpter 2006; Sumpter & Pratt 2009). In §3, we address the evolution of consensus decisions about movement destinations. We avoid the assumption of global overview by considering a self-organizing system.

3. EVOLUTION OF DECISION SHARING DURING MOVEMENTS BY SELF-ORGANIZING GROUPS
(a) Introduction
Many animal groups, in particular many economically important groups, are thought to be ‘self-organizing’ (see Couzin & Krause 2003 and Sumpter 2006 for reviews), in the sense that (i) pattern at the global level (e.g. group movement direction, group movement speed, group composition, group formation, etc.) emerges solely from numerous local interactions among the individual group members (i.e. nobody is in overall ‘command’) and (ii) the local behavioural rules specifying the interactions among the group members are executed using only local information, without reference to the global pattern (i.e. nobody has a global ‘overview’ of the group; Camazine et al. 2003). Typical examples for self-organizing groups are swarms of honeybees or locusts, colonies of ants, large shoals of fish, large flocks of birds and large herds of ungulates. Here, we are investigating the evolution of shared and unshared decision making in such self-organizing groups during group travel.

Studying the evolution of shared decision making in self-organizing systems poses some logistic challenges. Self-organizing animal systems are often studied with the help of computer simulations. However, these simulations can require considerable computing power (e.g. Couzin et al. 2005). Linking self-organizing systems modelling directly with evolutionary simulations, in order to answer evolutionary questions, would therefore require prohibitively large computing resources. We avoid this problem by adopting a novel approach in which we combine two different modelling techniques, namely, self-organizing system modelling and game theory modelling.

First, we modify a self-organizing system model by Couzin et al. (2005) in order to enable individual group members to differ in their local behaviours. We then use self-organizing system simulations to determine group level outcome into fitness consequences for individual group members (Conradt & Roper 2003, 2007). In this manner, we determine fitness consequences for each individual group member, depending both on its own local behaviour and on that of all other group members. The resulting pay-off matrices to individuals can then be used in multiplayer game theory models to determine evolutionarily stable strategies (‘ESSs’; Maynard-Smith 1989). Because multiplayer game theory models are complex, we use computer simulations to determine the relevant ESSs. Here, we use this method for the first time to investigate the evolution of decisions about movement destinations in a self-organizing group containing three members.

(b) Methods
(i) Conflicts of interest
To create conflicts of interest between members, we assume that two of the three members (‘majority members’) prefer to move to one spatial target, while the third member (‘minority member’) prefers to move to another spatial target. The spatial targets are equidistant (500 spatial units) from the group’s starting point, but lie in opposite directions. ‘type 1’ animals prefer to move to target 1 and ‘type 2’ animals to target 2. Thus, a group consists either of two type 1 and one type 2 animal, or vice versa.

(ii) Self-organizing group movement simulation model
The model is based on one by Couzin et al. (2005). The mathematical details are given in appendix Aa(i). All members move simultaneously in discrete time steps. Each individual member moves at each time step according to behavioural rules that depend only on local information (‘local movement rules’), as follows.
Table 1. Standardized relative pay-off matrix* to focal animal depending on its own and the other group member’s assertiveness, and on majority/minority relationships within the group.

(a) focal player is minority type in group

<table>
<thead>
<tr>
<th>focal animal plays</th>
<th>one majority player plays</th>
<th>other majority player plays</th>
</tr>
</thead>
<tbody>
<tr>
<td>$o_{low}$</td>
<td>$o_{medium}$</td>
<td>$o_{high}$</td>
</tr>
<tr>
<td>$o_{low}$</td>
<td>$o_{medium}$</td>
<td>$o_{high}$</td>
</tr>
<tr>
<td>$o_{medium}$</td>
<td>$o_{medium}$</td>
<td>$o_{high}$</td>
</tr>
<tr>
<td>$o_{high}$</td>
<td>$o_{high}$</td>
<td></td>
</tr>
<tr>
<td>$-0.6$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
<tr>
<td>$-0.7$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
<tr>
<td>$0$</td>
<td>$0.4 \cdot B'_2 - B'_3$</td>
<td>$-0.2 \cdot B'_3$</td>
</tr>
<tr>
<td>$0$</td>
<td>$1$</td>
<td>$-1$</td>
</tr>
<tr>
<td>$0$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
</tbody>
</table>

(b) focal player is majority type in group

<table>
<thead>
<tr>
<th>focal animal plays</th>
<th>other majority player plays</th>
<th>minority player plays</th>
</tr>
</thead>
<tbody>
<tr>
<td>$o_{low}$</td>
<td>$o_{medium}$</td>
<td>$o_{high}$</td>
</tr>
<tr>
<td>$o_{medium}$</td>
<td>$o_{medium}$</td>
<td>$o_{high}$</td>
</tr>
<tr>
<td>$o_{high}$</td>
<td>$o_{high}$</td>
<td></td>
</tr>
<tr>
<td>$-0.4$</td>
<td>$-0.9$</td>
<td>$-1$</td>
</tr>
<tr>
<td>$-0.4$</td>
<td>$0.6 \cdot B'_2 - B'_3$</td>
<td>$0.2 \cdot (B'_2 - B'_3) - 1$</td>
</tr>
<tr>
<td>$0$</td>
<td>$0$</td>
<td></td>
</tr>
<tr>
<td>$0$</td>
<td>$0.2 \cdot (B'_2 - B'_3) - 1$</td>
<td>$0$</td>
</tr>
<tr>
<td>$0$</td>
<td>$B'_2 - B'_3$</td>
<td>$0$</td>
</tr>
<tr>
<td>$0$</td>
<td>$B'_2 - B'_3$</td>
<td></td>
</tr>
</tbody>
</table>

*Note that if all animals within a group preferred the same target, the group always remained cohesive and moved towards the common preferred target; thus, for the evolution of behaviours, only those groups were relevant in which there were conflicts of interest between group members, and only those combinations are shown in the table.

The highest priority for an individual at any time is to avoid collision by moving away from neighbours that come too close. If (and only if) there is no immediate danger of collision, an individual has the following two aims: (i) to keep cohesion with other group members and (ii) to move towards its own preferred spatial target, as follows. In order to maintain group cohesion, the individual is attracted to neighbours within a given spatial range, so as to move towards them and align travel direction with them (‘social attraction’). However, the resulting ‘social attraction direction’ might differ from the direction towards the individual’s preferred spatial target. Therefore, the individual ‘balances’ attraction to its preferred spatial target against social attraction to neighbours with its own individual factor $o_i$ (‘degree of assertiveness’). In other words, it decides to move in the following direction: movement direction = social attraction direction + $o_i$-preferred target direction.

\[
\text{(3.1)}
\]

It follows that, if $o_i$ is large (and the individual is highly assertive), the individual decides to move at each time step predominantly in the direction of its preferred target; if $o_i$ is small (and the individual is of low assertiveness), it decides to move predominantly towards, and align with, neighbours; and if $o_i$ is intermediate (medium assertiveness), the individual compromises between its own target preference and moving towards, and aligning with, neighbours. Should the individual have no neighbours within the given spatial range, the individual has, at least for that time step, lost contact with other group members and cannot move towards, or align with, neighbours. Thus, it simply moves towards its own preferred spatial target (see above). According to these local behavioural rules, at each time step, for each individual, the next movement step direction was calculated (constrained by a maximum turning angle and including a random movement error; see appendix Aa(ii) for details), and then all individuals simultaneously moved in their new direction with a constant speed.

(iii) Variation in individual behaviours

The degree of assertiveness $o_i$ determines an individual’s readiness to compromise (see above). Therefore, it is the behavioural parameter that is most likely to influence the sharing/non-sharing of consensus decisions and the fitness of individual group members (Conradt et al. in press). Hence, we varied $o_i$ between individual members, such that they could exhibit low ($o_{low} = 0.05$), medium ($o_{medium} = 0.5$) or high ($o_{high} = 5.0$; see Couzin et al. 2005 for a biologically meaningful range of values for $o_i$) assertiveness. We consider all possible combinations $o_{12}, o_{13}, o_{23}$ (with $o_{12}, o_{13}, o_{23} \in \{o_{low}, o_{medium}, o_{high}\}$) for three members (i.e. 18 biologically different sets of behaviours; see also table 1). For each set of behaviours, we ran 500 simulations since, in pilot runs, results had stabilized to the required accuracy well below this number of replicates.

Phil. Trans. R. Soc. B (2009)
Relative gains

Therefore, in order to keep differences in relative benefits for different behavioural parameters to a minimum, we expressed the relative gains for each combination of behaviours. We averaged those gains over the 500 replicated simulations to be heading towards either target.

(v) Translating the results into individual fitness and pay-off matrices

We determined for each group member at the end of each simulation (i) the remaining number of members in their group and (ii) which target they were moving to (see above). This information can be translated into individual benefits, as follows. Firstly, social animals usually benefit by being in a group, and these ‘grouping benefits’ depend on group size (Krause & Ruxton 2002). We can, thus, determine the grouping benefits to an individual at the end of a simulation, depending on how many other group members it was still together with (solitary individuals gain no grouping benefits).

Secondly, individuals are likely to gain from moving towards their preferred spatial target and from moving towards the target preferred by other group members, and more so from the former than the latter (e.g. Conradt 1998; Ruckstuhl & Neuhaus 2000; Conradt & Roper 2003, 2007). The difference in fitness gains between moving towards a non-preferred target versus a preferred target constitutes a ‘consensus cost’ (Conradt & Roper 2003, 2005). Therefore, from the final position of individuals, we could determine what moving-to-preferred/non-preferred-target benefits an individual had gained, and whether an individual had incurred a consensus cost, or not. In this manner, we could, for each combination of behaviours, determine the net gains (grouping benefits + spatial target benefits) for each group member at the end of each simulation. We averaged those gains over the 500 replicated simulations for each combination of behaviours.

For the purpose of the game theory model, only differences in relative benefits for different behavioural strategies are relevant (rather than absolute benefits; Maynard-Smith 1989). Therefore, in order to keep parameters to a minimum, we expressed the relative benefits to individuals in each combination of behaviours in terms of grouping benefits relative to potential consensus costs. That is, $B_i$ and $B_j$ were the grouping benefits of remaining in a group with three and two members, respectively, relative to the potential consensus costs (i.e. the difference in benefits between moving to preferred versus non-preferred spatial target). Based on these two parameters, we could calculate the standardized relative pay-off matrix for majority and minority group members depending on their own behaviour and that of other group members (table 1). This pay-off matrix is the basis for the game theory model that follows below.

(vi) Game theory model

We make the conservative assumption that individuals in self-organizing groups do not have information about whether they are majority or minority types (Couzin et al. 2005). Thus, individuals behave independently of whether they are in a majority or minority. A focal individual can play the pure strategies $o_{low}, o_{medium}$ or $o_{high}$. It can also play any combination of these pure strategies. Assume that $r_{low}, r_{medium}$ and $r_{high}$ are the probabilities that the focal plays $o_{low}, o_{medium}$ or $o_{high}$ respectively. All possible complex behaviour strategies for a focal can, thus, be described as $r_{low}, r_{medium}, r_{high}$ (with $0 \leq r_{low} \leq 1$; $0 \leq r_{medium} \leq 1-r_{low}$ and $r_{high} = 1-r_{low}-r_{medium}$). Individuals drawn randomly from the population can also play the pure strategies $o_{low}, o_{medium}$ or $o_{high}$ or complex combinations. Assume that $p_{low}, p_{medium}$ and $q_{low}$ are the respective probabilities with which individuals that are drawn randomly from the population, and that have the same ($p_{low}$) or opposite ($q_{low}$) target preference as the focal, play $o_{low}, o_{medium}$ or $o_{high}$ respectively. Thus, all possible complex behaviour strategies in the population can be described as $p_{low}, p_{medium}, p_{high}$ and $q_{low}, q_{medium}, q_{high}$, as follows:

\[
\begin{align*}
\text{Relative gains} & \left\{ \begin{array}{c} r_{low}, r_{medium}, r_{high} \end{array} \right\} = r_{low} \left\{ (1-t)^2 \left[ -1 + 0.4q_{low}^2 + 2t(1-t) \right] \left( p_{low}(-0.4q_{low} - 0.9q_{medium} - q_{high}) \right) + p_{medium}(-0.4q_{medium} - q_{high}) + p_{high}(q_{high}(-0.4 + B_2 - B_3)) \right\} \\
& + r_{medium} \left\{ (1-t)^2 \left[ -1 + 0.9q_{low}^2 + 0.6q_{low}q_{medium} \right] + 2t(1-t) \left( p_{low}(-0.1q_{low} - 0.4q_{medium} - q_{high}) \right) + p_{medium}q_{high}(0.2(B_2 - B_3) - 1) + p_{high}q_{high}(B_2 - B_3) \right\} \\
& + r_{high} \left\{ (1-t)^2 \left[ (q_{high}(2 - q_{high}) + 0.2q_{medium}^2)(-B_3) + 0.8q_{low}q_{high}B_2 \right] \right\},
\end{align*}
\]

Conflicts of interest

L. Conradt & T. J. Roper 811

Phil. Trans. R. Soc. B (2009)
where \( t \) is the proportion of individuals in the population, which have the same target preference as the focal. We assumed that these relative gains translate into fitness consequences to individual players (Maynard-Smith 1989) and, thus, that \( q_{\text{low}} \), \( q_{\text{medium}} \), \( q_{\text{high}} \), and \( p_{\text{low}} \), \( p_{\text{medium}} \), \( p_{\text{high}} \) evolve accordingly (see also appendix A). We have further assumed that the proportion of different types of members \( t \) does not evolve. The rationale for this is as follows. Group members usually differ with respect to target preference because they are of different sex, age, size or physiological state (Conradt & Roper 2007), so \( t \) is commonly determined by factors (e.g. reproduction) that are independent of group decision making. However, this is not invariably so (see Rands et al. 2003) and, by covering a wide parameter range for \( t \) (0.1–0.9), we will be able to assess the potential effect \( t \) could have had if it had been allowed to co-evolve.

(vii) Evolutionarily stable strategies

We assume that relative gains (see equation (3.2)) translate into fitness consequences to individuals, and that the system evolves until it either reaches an evolutionarily stable strategy (Maynard-Smith 1989) or starts to oscillate. In order to find the ESSs in the system, we used a combination of analytical and simulation methods, as follows. Assume target type 1 is the focal target type, and target type 2 the other target type. First, we analytically determined stability conditions for pure strategies (i.e. \( p_{\text{low}} = 1 \), \( p_{\text{medium}} = 1 \) or \( p_{\text{high}} = 1 \); and \( q_{\text{low}} = 1 \), \( q_{\text{medium}} = 1 \) or \( q_{\text{high}} = 1 \); see appendix A(b(i))). The biological effect of the emerging pure ESSs (e.g. whether certain individuals or the majority led the group, and whether the group fragmented or not) can be derived from the self-organizing system simulations, given the pure strategies of both types (table 2). To examine the system for more complex ESSs, we used simulations. Starting from different ‘starting population strategies’, we simulated the evolution of population strategy, based on the relative gains given in equation (3.2) (assuming that these gains translate into fitness benefits; Maynard-Smith 1989), until there was either no further change in population strategy (i.e. we had reached an ESS), or the system began to oscillate. The details for these simulations are given in appendix A(b(ii)). By closely covering the whole parameter range for potential starting population strategies, we made sure we detected all potential ESSs in the system.

(c) Results

All of the ESSs that emerged were ‘pure’ ESSs, i.e. they consisted of a single pure strategy rather than an equilibrium mix of pure strategies (figures 1 and 2, table 2). There were six possible ESSs (table 2). Their biological consequence for group decisions ranged from (i) complete segregation of different target types, via (ii) partial segregation, and (iii) equally shared consensus decision making, to (iv) consensus decision making by individuals of only one of the two target types. We examine the ESSs, assembled by their biological consequences, in detail, in the following.

(i) Complete segregation of target types

If the benefits of being in a group of three are lower than the potential consensus costs (i.e. if \( B'_{3} < 1 \); figure 1, black areas) for both target types, the only ESS in the system is: ‘both target types always play \( \omega_{\text{high}} \)’ (table 2). That is, none of the group members should compromise and incur consensus costs in order to maintain group cohesion (and gain grouping benefits that do not outweigh the consensus costs). This ESS leads to complete segregation of different target types into separate groups. In the rest of the results section, we only consider what happens if grouping benefits relative to potential consensus costs are high enough not to automatically lead to complete segregation (i.e. if it is \( B'_{3} > 1 \) for at least one target type).

(ii) Partial segregation of target types

If group size three is above optimal group size (i.e. if \( B'_{3} < B'_{2} \)), and grouping benefits of being in a group of three are less than five times as high as potential consensus costs for at least one of the two target types (i.e. if \( B'_{3} < 5 \)), three possible situations can arise, as follows (see table 2 for criteria for each case). Firstly, no ESS exists and both types oscillate between playing \( \omega_{\text{high}} \) and \( \omega_{\text{medium}} \) (figure 1, horizontally striped areas). Secondly, one ESS exists: ‘more common type/type medium’ (figure 1, light grey areas). Thirdly, two ESSs exist: (i) ‘type 1 always plays \( \omega_{\text{high}} \) and type 2 always \( \omega_{\text{medium}} \)’ and (ii) ‘type 2 always plays \( \omega_{\text{high}} \) and type 1 always \( \omega_{\text{medium}} \)’ (figure 1a, light grey area). The biological effect is the same in all three situations, namely members will not always reach consensus and at least some group fragmentation will occur in mixed-type groups. This will lead to at least partial segregation of the two target types into separate groups.

(iii) Equally shared consensus decisions

If, and only if, grouping benefits of being in a group of three are more than five times as high as potential consensus costs for both target types (i.e. if \( B'_{3} > 5 \)), the following strategy was always an ESS: ‘both target types always play \( \omega_{\text{medium}} \)’. The biological effect of this ESS is equally shared consensus decision making (see table 2; cf. Couzin et al. 2005). Note, however, that only disjoint parameter values for \( \omega \) could be tested here, and that ‘both target types always play \( \omega_{\text{medium}} \)’ might cease to be an ESS in a continuous parameter space for \( \omega \). It would then be likely to be replaced by another ‘equal-sharing’ ESS with a slightly higher common \( \omega \) value and slightly different stability criteria, while qualitative results would be similar (Conradt & Roper 2007).

If the minimum condition \( B'_{3} > 5 \) was met for both target types, and group size three was above optimal group size (i.e. if \( B'_{3} < B'_{4} \)), both target types always play \( \omega_{\text{medium}} \) was the only ESS in the system, and equal sharing of consensus decisions always evolved (figure 1, spotted area). If group size three was not above optimal group size (i.e. if \( B'_{3} > B'_{3} \); figure 1, white areas), two further ESSs could exist: ‘type 1 always plays \( \omega_{\text{low}} \) and type 2 always \( \omega_{\text{high}} \)’ (which would mean that type 2...
Table 2. Evolutionary stability criteria for pure strategies of either target type.

<table>
<thead>
<tr>
<th>Strategy of type 1 individuals</th>
<th>Strategy of type 2 individuals</th>
<th>Stability criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Play always</td>
<td>Play always</td>
<td>Never stable</td>
</tr>
<tr>
<td>ESS if $E_1$ ($type 1$) &gt; $E_2$ ($type 2$) &gt; $E_{mix}$ ($medium$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ii) $type 2$ always play low</td>
<td>(i) $type 1$ always play low</td>
<td></td>
</tr>
<tr>
<td>$p_{low}$ ($low$)</td>
<td>$q_{low}$ ($low$)</td>
<td></td>
</tr>
<tr>
<td>and $E_1$ ($type 1$) &gt; $E_2$ ($type 2$) &gt; $E_{mix}$ ($medium$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(iii) $type 1$ leads all decisions; majority leads equally shared decision; some group fragmentation, hence, partial segregation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(iv) One target type leads consensus decisions</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We consider only situations without complete or partial segregation (see above and table 2 for criteria). If group size three is not above optimal group size (i.e. if $B_1 > B_2$), the following two strategies are potential ESSs: (i) ‘$type 1$ always plays $\omega_{high}$ and type 2 always $\omega_{low}$’ and (ii) ‘$type 2$ always plays $\omega_{high}$ and type 1 always $\omega_{low}$’. Either of these ESSs results in decisions being made by individuals of only one target type. In particular, the first ESS means that target type 1 individuals lead all consensus decisions (table 2; figure 1, dark grey and vertically striped areas), the second ESS that target type 2 individuals lead all consensus decisions. Since individuals of the leading target type always play $\omega_{high}$, they do not compromise with other group members but behave domineeringly. Decisions are no longer equally shared, and the decision is not necessarily made by members of the target type that are in the majority within the group. If only one member in a group is of the ‘leading’ target type, the decision will, strictly speaking, even be ‘unshared’.

To which ESS the system evolves, depends on the starting point of the system, the proportion $t$ of type 1 in the population and on the grouping benefits relative to potential consensus costs (i.e. $B_1$ and $B_2$). If the ratio of grouping benefits to potential consensus costs is similar for individuals of both types (i.e. if $B_1$ ($type 1$) = $B_2$ ($type 2$) and if $B_1$ ($type 1$) = $B_2$ ($type 2$)), and if $t$ is small, the ESSs ‘$type 1$ individuals lead all consensus decisions’ and ‘$type 2$ individuals lead all consensus decisions’ are equally likely to evolve (figure 1a). Otherwise, if grouping benefits are small relative to potential consensus costs (i.e. $B_1$ is small; figure 1, dark grey area; see table 2 for threshold), only that ESS can evolve in which those individuals lead all decisions, which are of the more common type (figure 1b) and/or of the type with lower grouping benefits relative to potential consensus costs (figure 1c–e). As grouping benefits relative to potential consensus costs increase (i.e. as $B_1$ increases), the system can evolve also to ‘leading by the other type’ (figure 1, vertically striped areas). If grouping benefits of being in a group of three exceed five times the potential consensus costs (i.e. if $B_1 > 5$) for both target types, a third ESS exists that leads to equal sharing of consensus decisions (see section above). If relative grouping benefits increase further (i.e. $B_1$ increases further, and in particular if $B_1$ increases), the system is more and more likely to evolve to equal sharing of
Figure 1. ESSs depending on the grouping benefits to type 1 of being in a group of three ($B'_x$, x-axis) or two ($B'_y$, y-axis), respectively, relative to the potential consensus costs; (a) both types with same grouping benefits to consensus cost ratio (i.e. $B'_1$ (type 1) = $B'_2$ (type 2); $B'_3$ (type 1) = $B'_2$ (type 2)), $t = 0.5$; (b) both types with same grouping benefits to consensus cost ratio, $t = 0.1/0.9$; (c) type 1 with four times higher grouping benefits to consensus cost ratio (i.e. $B'_1$ (type 1) = 4 $B'_2$ (type 2); $B'_3$ (type 1) = 4 $B'_2$ (type 2)), $t = 0.1$; (d) type 1 with four times higher grouping benefits to consensus cost ratio, $t = 0.5$; (e) type 1 with four times higher grouping benefits to consensus cost ratio, $t = 0.9$ (black areas, 1 ESS, all individuals play always $b_{high}$, resulting in complete segregation; horizontally striped areas, no ESS, individuals oscillate between playing $b_{medium}$ and $b_{high}$, resulting in varying leading and partial segregation; light grey areas, 1 (or 2) ESSs, the type with the highest potential consensus costs or which is more common plays always $b_{high}$, the other type plays always $b_{medium}$, resulting partially in leading by first type and partially in segregation; dark grey areas, 1 ESS, the type with the highest potential consensus costs or which is more common plays always $b_{high}$, the other type plays always $b_{low}$ resulting in leading of decisions by first type; vertically striped areas, 2 ESSs, one type always plays $b_{high}$, the other type always $b_{low}$ resulting in ‘leading by type 1’ or ‘leading by type 2’; spotted areas, 1 ESS, all types play always $b_{medium}$, resulting in equally shared decision making; white areas, 3 ESSs, ‘leading by type 1’, ‘leading by type 2’ and ‘equally shared decision making’).
Roper 2007) would go completely extinct. Neither target type (e.g. sex, age or size class; Conradt & Roper 2003) would have led to great changes, as collective decisions are made which maximize total group benefits. This implies that an evolving consensus costs increased, equal sharing of decisions. The reason for this apparent contradiction lies in the different nature of decisions about activity synchronization (which were the subject of Conradt & Roper 2007) and decisions about movement destinations. In decisions about the timing of activities, compromises that can be reached in an equally shared manner (e.g. the average of median of preferred timings by group members) usually lower the average consensus costs to individual group members (Conradt & Roper 2003). This is generally true if the modality which the group decides upon (e.g. timing) is continuous with respect to benefits/costs. In such a case, a compromise can be viable. To give an example: if you want to go to a restaurant at 19.00 hours, and I want to go at 20.00 hours, going to 19.30 hours might mean that both of us do not get inconveniently too much. To compromise (and share a decision) on such a continuous modality will be particularly crucial if potential consensus costs are high and grouping benefits marginal, since group members should only stay in a group if grouping benefits exceed consensus costs. Thus, in such situations, one expects either equal or shared decisions or group fragmentation. Group decisions about timing, while not always, will usually fall into this category.

By contrast, if the modality which the group decides upon (e.g. spatial targets) is disjunct with respect to benefits/costs, then a compromise is unlikely to be viable. To give an example: if you are hungry and want to go to a restaurant, and I need petrol and want to go to a garage, compromising and going halfway between the restaurant and the garage (where we could neither get food nor petrol) does not make sense for either of us. While not all decisions on spatial targets are necessarily ‘disjunct decisions’ in this way, the majority are, most notably in patchy landscapes. Such circumstances usually require ‘hard’ decisions, of an ‘either/or’ kind, to be made. Thus, groups in the kind of situation that we have considered here have only three viable options: move to target 1, move to target 2 or fragment. If grouping benefits are large relative to consensus costs, no group member wishes to risk group fragmentation, so being too assertive with respect to your own preference is risky and therefore disadvantageous. On the other hand, if nobody is highly assertive it pays to be somewhat assertive, because otherwise you would never move to your own preferred target. Consequently, evolution favours a medium level of assertiveness in all individuals, which leads in a self-organizing system to equally shared decision making (Couzin et al. 2005). By contrast, if grouping benefits are low relative to consensus costs, risking group fragmentation rather than moving towards a less preferred target becomes an evolutionary option. Thus, it will pay some individuals, in particular those which would face higher consensus costs if moving towards the less preferred target (or those which stand to gain less grouping benefits), to be highly assertive. As individuals of one target type evolve to be highly assertive, selection might then favour individuals of the other target type to be unassertive in order to avoid group fragmentation. The result is that ‘leading by one type’ evolves. If grouping benefits drop further,
individuals with different preferences segregate into different subgroups.

An interesting implication of this is that if animals make simultaneous group decisions about continuous and disjunct modalities (e.g. about space and time), they should apply different strategies to the two different aspects of the decision. How the details of combined space-and-time decisions could look requires further modelling.

The model proves that equal sharing of decisions is not a foregone conclusion, even in global group behaviours (Camazine et al. 2003). On the contrary, we show that leading by one type can evolve, even if group members cannot communicate, and consequently have no knowledge about, the preferences of other group members. The model offers testable predictions about when equal sharing or leading by one type should occur in group decisions about movement destinations.

The present model is for computational reasons restricted to groups of three. However, its results provide initial insight into an important type of group decision (see Couzin & Krause 2003 for a review). Its most important qualitative result (i.e. the principal difference between decisions in continuous and disjunct modalities, and thus, usually between timing and spatial decisions) is likely to hold for larger group sizes, for the reasons given above. Whether other results also hold requires further investigation. As computer speeds increase, our approach will become feasible for studies on larger groups.

L.C. was supported by a Royal Society University Research Fellowship, and would like to thank the Royal Society particularly for its support during maternity leave and with respect to part-time working.

APPENDIX A

(a) Details of the self-organizing group movement simulation model

(i) Individual local behaviour rules

Collision avoidance. For each individual, the highest priority is to avoid collision. Thus, if there are neighbours within a collision range $\alpha$ ($\alpha=2$ for adequacy of parameter choice see Couzin et al. 2005), the individual simply tries to turn away from those neighbours. Thus, if $N_{\text{in}}(t)>0$, then

$$d_i(t) = -\frac{1}{N_{\text{in}}(t)} \sum_{j=1}^{N_{\text{in}}(t)} \frac{r_{ij}(t) - c_i(t)}{|r_{ij}(t) - c_i(t)|}, \quad (A1)$$

where $N_{\text{in}}(t)$ is the number of neighbours within range $\alpha$; $d_i(t)$ is the preliminary movement direction; and $c_i(t)$ is the position vector, for individual $i$ at time $t$.

Balance between social attraction and preferred target direction. If there are no neighbours within range $\alpha$ (i.e. if $N_{\text{in}}(t)=0$), the individual’s movement is based on its desire to maintain group cohesion as well as to move towards its preferred target, as follows. In order to maintain group cohesion, the individual is attracted to neighbours within a range $\rho$ ($\rho=30$; for adequacy of parameter choice see Couzin et al. 2005) so as to move towards them and align travel direction with them (‘social attraction’). However, the resulting ‘social attraction vector’ might differ from the vector towards the individual’s preferred spatial target. Therefore, the individual ‘balances’ attraction to its preferred target against social attraction to neighbours with its own individual factor $\omega_i$ (‘degree of assertiveness’). That is, an individual tries to move in the direction of social attraction $+\omega_i$-preferred target direction (where preferred target direction is the vector between the individual’s position and the preferred target position, scaled to unit length). Thus, if $N_{\text{in}}(t)=0$ and $N_{\text{in}}(t)>0$, then

$$d_i(t) = \frac{1}{4N_{\text{in}}(t)} \sum_{j=1}^{N_{\text{in}}(t)} \frac{v_j(t)}{|c_j(t) - c_i(t)|} + \frac{1}{4N_{\text{in}}(t)} \sum_{j=1}^{N_{\text{in}}(t)} \frac{v_j(t)}{|c_j(t) - c_i(t)|} + \omega_i \frac{g_i - c_i(t)}{2|g_i - c_i(t)|}, \quad (A2)$$

where $N_{\text{in}}(t)$ is the number of neighbours within range $\rho$; $v_j(t)$ is the unit vector of movement direction, for individual $j$ at time step $t$; and $g_i$ is the position vector of the preferred target of individual $i$. The first term describes social attraction to neighbours, the second term alignment with neighbours (first and second term together constitute the social attraction vector), and the third term attraction to the preferred goal.

It follows that, if $\omega_i$ is large (and the individual is highly assertive), the individual tries to move at each time step predominantly in the direction of its preferred target; if $\omega_i$ is small (and the individual is of low assertiveness), it predominantly tries to move towards, and align with, neighbours; and if $\omega_i$ is intermediate (medium assertiveness), the individual compromises between its own target preference and moving towards, and aligning with, neighbours.

Absence of neighbours. If there are no neighbours within ranges $\alpha$ or $\rho$, the individual always tries to move in the direction of its preferred target. Thus, if $N_{\text{in}}(t) = 0$ and $N_{\text{in}}(t) = 0$, then

$$d_i(t) = \frac{g_i - c_i(t)}{|g_i - c_i(t)|}. \quad (A3)$$

Note that none of these assumptions regarding local behavioural rules requires a global overview of the group, or a view of other group members further away than distance $\rho$.

(ii) Self-organizing system simulations

Individual group members start with a random movement direction and a random position within a cohesive group. From the preliminary movement direction $d_i(t)$ (equations (A1)–(A3), above), we derive the new movement direction for each individual at the next time step by adding an individual random directional error to $d_i(t)$ (with standard deviation $\sigma=0.01$ radians; Couzin et al. 2005), and limiting the change in movement direction to a maximum turning angle of $\theta \Delta t$ ($\theta \Delta t=0.2$ radians; see Couzin et al. 2005 for details and justification of parameter choice). The new position of each individual is obtained by moving it one step ahead of $\Delta t=0.1$ (equivalent to a constant speed of 1 spatial unit/time unit) in the new movement direction.
(b) Game theory model

(i) Criteria for the increases in the population of different behavioural strategies

We determine the criteria under which different pure behavioural strategies should increase or decrease in the population, using the expected relative gains given in equation (3.2). Equation (3.2) can be written shorter as

\[ \text{Relative gains} = \{p_{\text{low}}, r_{\text{medium}}, r_{\text{high}}\}, \]

\[ = r_{\text{low}}A_t^{\text{low}} + r_{\text{medium}}B_t^{\text{medium}} + r_{\text{high}}C_t^{\text{high}}, \]

These relative gains peak with \( r_{\text{low}} = 1 \) if \( A_t^{\text{low}} \) is relative large (which means that it is of advantage to play \( p_{\text{low}} \)). Similarly, they peak with \( r_{\text{medium}} = 1 \) if \( B_t^{\text{medium}} \) is relative large (which means that it is of advantage to play \( q_{\text{low}} \)). Finally, they peak with \( r_{\text{high}} = 1 \) if \( C_t^{\text{high}} \) is relative large (which means that it is of advantage to play \( q_{\text{high}} \)). We assume that relative gains translate into relative fitness increases (Maynard-Smith 1989). It therefore follows that \( p_{\text{low}} \) (i.e. the probability with which individual’s focal’s target preference play \( \omega_{\text{low}} \)) increases in the population relative to other strategies, if relative gains reach a maximum for \( r_{\text{low}} = 1 \). Similarly, \( p_{\text{medium}} \) increases in the population if relative gains reach a maximum for \( r_{\text{medium}} = 1 \). Finally, \( p_{\text{high}} \) increases in the population if relative gains reach a maximum for \( r_{\text{high}} = 1 \). In particular, \( p_{\text{low}} \) increases in the population if

\[ A_t^{\text{low}} > B_t^{\text{low}}, \]

and

\[ C_t^{\text{low}} > B_t^{\text{low}}. \]

(A 5a)

\[ A_t^{\text{medium}} > B_t^{\text{medium}}, \]

and

\[ C_t^{\text{medium}} > B_t^{\text{medium}}. \]

(A 5b)

\[ p_{\text{medium}} \] increases in the population if

\[ B_t^{\text{medium}} > A_t^{\text{medium}}. \]

(A 6a)

\[ B_t^{\text{high}} > C_t^{\text{high}}, \]

and \( p_{\text{high}} \) increases in the population if

\[ C_t^{\text{high}} > A_t^{\text{high}}. \]

(A 7a)

and

\[ C_t^{\text{low}} > A_t^{\text{low}}. \]

(A 7b)

Of course, the analogue is true for the members in the population of the opposite target preference. Thus, \( q_{\text{low}} \) increases in the population if

\[ A_t^{\text{low}} > B_t^{\text{low}} \]

and

\[ C_t^{\text{low}} > B_t^{\text{low}}. \]

(A 8a)

\[ p_{\text{medium}} \] increases in the population if

\[ B_t^{\text{medium}} > A_t^{\text{medium}} \]

and

\[ C_t^{\text{medium}} > B_t^{\text{medium}}. \]

(A 8b)

\[ q_{\text{high}} \] increases in the population if

\[ C_t^{\text{high}} > B_t^{\text{high}} \]

and

\[ p_{\text{high}} \] increases in the population if

\[ A_t^{\text{high}} > B_t^{\text{high}}. \]

(A 9a)
Our starting population strategies covered the whole parameter space (in steps of 0.1) and used all biologically possible combinations of $p_{low} \in \{0, 0.1, 0.2, \ldots, 1\}$; $p_{medium} \in \{0, 0.1, 0.2, \ldots, 1-p_{low}\}$; $q_{low} \in \{0, 0.1, 0.2, \ldots, 1\}$; and $q_{medium} \in \{0, 0.1, 0.2, \ldots, 1-q_{low}\}$. Thus, by closely covering the whole range of potential starting population strategies, and by using small increases at each simulation step, we made sure that we detected all potential ESSs in each system. The examined parameter range for proportion $t$ was 0.1, 0.5 and 0.9; and that for relative grouping benefits $B_2^1$ (type 1), $B_2^2$ (type 1) and $B_2^3$ (type 2): 0.5–25 (in steps small enough to cover all potentially different cases arising from the stability conditions in table 2). It was assumed that the ratio of grouping benefits relative to potential consensus costs ($B_2^1$ and $B_2^3$) were either the same for individuals that preferred targets 1 and 2 (i.e. $B_2^1$ (type 1)=$B_2^2$ (type 2) and $B_2^3$ (type 1)=$B_2^3$ (type 2)), or they were four times higher for individuals that preferred target 1 than for those that preferred target 2 (i.e. $B_2^1$ (type 1)=$4 \cdot B_2^2$ (type 2) and $B_2^3$ (type 1)=$4 \cdot B_2^3$ (type 2)).

REFERENCES


