Published in Journal of Evolutionary Biology 21, issue 3, 405–420, 2008
which should be used for any reference to this work

Distinguishing four fundamental approaches to the evolution of helping

R. BSHARY & R. BERGMÜLLER

Department of Biology, University of Neuchâtel, Neuchâtel, Switzerland

Abstract

The evolution and stability of helping behaviour has attracted great research efforts across disciplines. However, the field is also characterized by a great confusion over terminology and a number of disagreements, often between disciplines but also along taxonomic boundaries. In an attempt to clarify several issues, we identify four distinct research fields concerning the evolution of helping: (1) basic social evolution theory that studies helping within the framework of Hamilton's inclusive fitness concept, i.e. direct and indirect benefits, (2) an ecological approach that identifies settings that promote life histories or interaction patterns that favour unconditional cooperative and altruistic behaviour, e.g. conditions that lead to interdependency or interactions among kin, (3) the game theoretic approach that identifies strategies that provide feedback and control mechanisms (protecting from cheaters) favouring cooperative behaviour (e.g. pseudo-reciprocity, reciprocity), and (4) the social scientists' approach that particularly emphasizes the special cognitive requirements necessary for human cooperative strategies. The four fields differ with respect to the 'mechanisms' and the 'conditions' favouring helping they investigate. Other major differences concern a focus on either the life-time fitness consequences or the immediate payoff consequences of behaviour, and whether the behaviour of an individual or a whole interaction is considered. We suggest that distinguishing between these four separate fields and their complementary approaches will reduce misunderstandings, facilitating further integration of concepts within and across disciplines.

Keywords

altruism; cognition; control mechanism; cooperation; ecology; kin selection; life histories; strategies.

Introduction

The evolution and maintenance of helping, which we define as a behaviour that increases the direct fitness of another individual (Lehmann & Keller, 2006), remains a key puzzle that needs to be explained within the framework of evolutionary theory. Why should an individual conduct a behaviour that provides benefits for others if selection favours individuals that maximize their own fitness? Several reviews and target papers have been published on this subject in recent years (Hammerstein, 2003a; Bshary & Bronstein, 2004; Sachs *et al.*, 2004; Lehmann & Keller, 2006; Noë, 2006; Nowak, 2006; West *et al.*, 2007; Bergmüller *et al.*, 2007a). The field of cooperation has become a truly interdisciplinary medley of research as evident from a recent edited book (Hammerstein, 2003a). Whereas this is exciting, problems arise when scientists from different disciplines or even subdisciplines try to communicate with each other for two main reasons: (1) they use different terminology and (2) they have different traditions regarding methodology (Noë, 2006), reflecting the different kinds of question that are addressed in the various fields.

Correspondence: Redouan Bshary, Department of Biology, University of Neuchâtel, Emile-Argand 11, Case Postale 158, 2009 Neuchâtel, Switzerland.
Tel.: +41 32 7183005; fax: +41 32 7183001;
e-mail: redouan.bshary@unine.ch

West *et al.* (2007) show the pitfalls of inconsistent definitions and bravely propose a coherent framework for terminology. The more fundamental historical problem, however, is the diversity of methods and conceptual approaches that are currently used. These differences are often hidden because our everyday use of language suggests that the same question is asked. Tinbergen (1963) was the first to point out that a seemingly simple question like 'why does a starling sing?' may be answered in four fundamentally different ways. One may try to understand the mechanisms or the ontogeny of the behaviour (proximate questions) or one may try to understand the adaptive value or the phylogeny of the behaviour (ultimate questions) (Mayr, 1961; West *et al.*, 2007). Only if one specifies which of these four approaches one is talking about it is possible to eliminate the most basic source of confusion.

We address the evolution and stability of cooperation in this article, so we ask an ultimate question about the adaptive value of behaviour, which is the key question for all 'evolutionary scientists' (evolutionary economists, anthropologists, psychologists and biologists). However, we will try to clarify that the current research on the general question about the adaptive value of helping can be subdivided in four more specific research topics, each of which approaches the issue from a different angle. Our distinction of four fields has partly been foreshadowed by two recent conceptual papers (Lehmann & Keller, 2006; West *et al.*, 2007) to which will refer repeatedly in the article. Similar to the '4 whys' in biology, the four research topics on the evolution and stability of helping must be kept separate to avoid fruitless discussions. The basis for some confusion can be attributed to terminology (West *et al.*, 2007; Bergmüller *et al.*, 2007a). In particular, all researchers talk about the 'conditions' and the 'mechanisms' that promote helping. These terms mean different things, however, depending on the specific research field (Table 1). The result is that researchers often believe they are addressing the same questions, while the actual theoretical models or the empirical data may in fact tackle very different issues. We propose that in order to avoid useless debates it is important to distinguish four major aspects of helping that are currently investigated, namely (1) basic social evolution theory, which explores the evolutionary pathways that select for helping. Helping only evolves under the *condition* that the actor's inclusive fitness is increased, while the *mechanism* is either an increase in direct or in indirect fitness. (2) The ecological approach still focuses on direct and indirect benefits as *mechanisms* for the evolution of helping, while the necessary *conditions* include life history parameters and social systems. (3) The game theoretic approach 'translates' ecological conditions into a game structure (*n* interactions, payoff matrix), which provides the *conditions* under which control *mechanisms* like reciprocity, punishment, etc., may ensure that any form of investment yields on average an increase in the actor's fitness. (4) The social scientists' approach identifies psychological and physiological *mechanisms* that promote helping in humans, while the *conditions* comprise moral values or the existence of specific brain structures.

We will develop the differences between the four fields in more detail later in this article, but we give a quick first illustration with the example of a human paying the bill for an unrelated individual in a restaurant. According to social evolution theory, this act of helping must increase the inclusive fitness of the actor, and in our example through direct fitness benefits. The ecological approach would specify that humans are long lived and social, and that there is therefore a high probability of repeated interactions and interdependency between the two individuals, which facilitates the evolution of helping through direct fitness benefits. A game theoretician would derive a payoff matrix to specify the costs and benefits of the act, calculate the probability that the two meet again in the future, and explore how the actor could possibly force reciprocation if it is not given freely. Finally, cognitive scientists would explore whether such giving is correlated with moral values that make the actor feel good about helping someone else, and whether this is achieved through stimulation of the reward centre in the forebrain. We hope this example helps clarifying that the four fields are indeed complementary and that they address different questions. As a consequence, we need a broad range of terms to capture the various questions of interest and at the same time to avoid that the same term has several meanings.

Table 1 The use of the terms 'conditions' and 'mechanisms' in four distinct approaches to the evolution and maintenance of helping behaviour.

	Evolutionary pathways	Ecological settings	Strategies	Social scientists' approach
Conditions	Increase in inclusive fitness	Overlapping generations, low migration, group living, etc.	*n* interactions, payoff matrix, body condition, etc.	Culture, moral, empathy, specific brain structure
Mechanisms	Direct or indirect fitness benefits	Direct or indirect fitness benefits	Punishment, reward, partner switching, termination of interaction, etc.	Psychological: guilt, pleasure, etc. Physiological: oxytocin, brain stimulation, etc.

Terminology

Whereas the goal of this paper is to highlight the importance of distinguishing four major research topics concerning the evolution of helping, we cannot escape the issue of terminology. Our feeling, based on a recent target article with 22 replies (special issue in Behavioural Processes) is that the field is too much grown and too diverse for any proposed terminology to become universally accepted. Therefore, it is very important that we give clear definitions for every term we use so that each reader can translate the content to her/his use of terminology. Our terminology attempts to fulfil three criteria. (1) The terminology should allow us to cover the research questions addressed in the four fields. (2) We should avoid using the same term in different research fields but with different meanings. (3) We should try to accommodate recent attempts to clarify the terminology.

We propose that in order to address all relevant issues on the evolution of helping, we need terms that cover four different aspects (without corresponding one to one to our four research fields). (1) A classification of social *behaviours* by their influence on the *lifetime direct fitness* of actor and recipient (keeping the recipient's behaviour constant). (2) A classification of social *interactions* by the impact of two players on each other's *lifetime direct fitness*. (3) A classification of social *behaviours* in the context of cooperation by their influence on the *immediate payoff* of actor and recipient (keeping the recipient's behaviour constant). (4) A classification of social *interactions* in the context of cooperation by the impact of two players on each other's *immediate payoff*.

In order to distinguish between behaviours of individuals and the outcome of interactions, we use verb forms for behaviours (following Hamilton, 1964, 1970) and nouns for interactions, the latter being of key importance in game theory (Dugatkin, 1997). We generate four 2×2 tables that capture the short-term or long-term outcomes of either social behaviours or social interactions and our corresponding terminology (Box 1). Below follows a brief reasoning for our terminology.

1. A classification of social *behaviours* by their influence on the *lifetime direct fitness* of actor and recipient (keeping the recipient's behaviour constant)

We need terms to formulate basic social evolution theory. West *et al.* (2007) discuss the current confusion in terminology in detail and propose a coherent framework for terms that describe the *average impact* of a social behaviour on the *direct fitness* of actor and recipient. We adapted their terminology to our purposes by transforming their terms, which were given as nouns, into verbs. We refer to behaviour that increases the direct fitness of both actor and recipient (+/+) as mutually beneficial behaviour. Altruistic behaviour reduces the direct fitness of the actor while increasing the direct fitness of the recipient (−/+). A selfish behaviour increases the direct fitness of the actor while

Box 1

How to explain the evolution of helping (−/+ or +/+): the necessary terminology.

1. A classification of social **behaviours** by their influence on **lifetime** direct fitness of actor and recipient (keeping the recipient's behaviour constant). Mutually beneficial behaviour and altruistic behaviour can be summarized as helping.

	Recipient	
	+	−
Actor		
+	Mutually beneficial behaviour	Selfish behaviour
−	Altruistic behaviour	Spiteful behaviour

2. A classification of social **interactions** by the impact of two players on each other's **lifetime** direct fitness

	Player 2	
	+	−
Player 1		
+	Cooperation (within species) Mutualism (between species)	Altruism/parasitism/predation
−	Altruism/parasitism/predation	Competition Spite

3. A classification of social **behaviours** in the context of cooperation by their influence on the **immediate** payoff of actor and recipient (keeping the recipient's behaviour constant)

	Recipient	
	+	−
Actor		
+	Self serving mutually beneficial behaviour	Cheating
−	Investing	Punishing

4. A classification of social **interactions** in the context of cooperation by the impact of two players on each other's **immediate** payoff

	Player 2	
	+	−
Player 1		
+	Mutual cooperation	Exploitation
−	Exploitation	Mutual defection

it decreases the direct fitness of the recipient (+/−) and spiteful behaviour decreases the direct fitness of both actor and recipient (−/−). Note that our use of the verb form is in line with Hamilton (1964, 1970) who also used verbs to describe social behaviours.

2. A classification of social *interactions* by the impact of two players on each other's *lifetime direct fitness*

While basic social evolution theory keeps the behaviour of recipients constant and hence focuses on the behaviour of the actor, each individual is an agent and it is therefore of interest to explore the impact of interaction partners on each other's *average* lifetime *direct fitness*. West *et al.* (2007) seem to agree in part with this distinction as they note that studies on interspecific mutualism describe the impact that each species has on the other (i.e. the result of the interaction) and acknowledge this question to be different from basic social evolution theory (West *et al.*, 2007, p. 4; right column). The terminology we use is very much in line with standard ecological literature (Begon *et al.*, 2005), which applies to both intraspecific and interspecific interactions. We define a mutual positive influence (+/+ outcome) on the direct fitness of interacting individuals as cooperation if the partners belong to the same species and as mutualism if the partners belong to different species (Bronstein, 2001; Bshary & Bronstein, 2004). A −/+ outcome of interactions is altruism if the outcome is due to 'mutual agreement'. Otherwise, a −/+ outcome may be due to parasitism or predation. Finally a mutually negative effect on each other (−/−) could either be due to mutual spite or due to competition.

3. A classification of social *behaviours* in the context of cooperation by their influence on the *immediate payoff* of actor and recipient (keeping the recipient's behaviour constant)

Whereas social evolution theory explores rather simple conditions that allow to precisely describing the conditions under which a behaviour is selected for, real life situations are usually variable, and so is the behaviour of individuals. Learning through positive or negative reinforcement may modify behaviour in virtually all animals (Wynne, 2001). Given that behaviour is often flexible and dependent on previous experience, we need terms that describe how a behaviour influences the immediate payoff of actor and recipient. If the behaviour has immediate positive effects for both actor and recipient, we term this a *self serving* mutually beneficial behaviour (Cant & Johnstone, 2006). If the actor benefits while the recipients looses, we call the behaviour cheating. If the actor has immediate costs and the recipient immediate gains, we call the behaviour investment. Finally, we call a behaviour that reduces the immediate payoffs for both actor and recipient 'punishment', following Clutton-Brock & Parker (1995). These authors note that punishment is 'temporarily spiteful' to emphasize the −/− description of the payoffs.

4. A classification of social *interactions* in the context of cooperation by the impact of two players on each other's *immediate payoff*

Behaviour is often embedded in conditional strategies, where current behaviour of self and of interacting partners, in combination with the payoff received, may influence future behaviour. An individual must be able to respond appropriately to the behaviour of others. In this context, the distinction between the action of a single individual and the interaction between two individuals is a fundamental aspect of game theoretic analyses (Dugatkin, 1997; Bergmüller *et al.*, 2007a). Certainly all of us will have had experience in investing in another individual in hope of a return on that investment which never materialized. In such situations we behaved cooperatively but the interaction was not cooperation. In conclusion, the payoff of each player depends on both its own behaviour and on how the other player behaves. We therefore need terms that describe the outcomes of this 2×2 matrix. We term the mutual increase in payoffs 'mutual cooperation', whereas a positive payoff for one player and a reduction in the payoff of the other is termed 'exploitation'. Finally, a mutually negative consequence on each other's payoff is termed 'mutual defection'.

As can been seen from the four 2×2 matrices in Box 1, we managed to avoid using any term twice except for the term 'cooperation', which we used for mutually beneficial outcomes of interactions both in the short term and with respect to average direct fitness consequences. However, we think that the double use of cooperation is not problematic as it seems logical that a mutual increase in immediate payoffs will also translate into a mutual increase in average direct fitness.

Surprisingly, we have not yet used the term 'cooperative behaviour' for our terminology in Box 1. In a way, this is an advantage because this term has been used in so many different ways in the literature. This seems to be due in part to the fact that some scientists study the short-term consequences of behaviour in the form of payoff matrices, while others are interested in the average fitness consequences. Keeping this distinction in mind, we define the term 'cooperative behaviour' in the short-term sense as a behaviour that is either self-serving mutually beneficial or an investment (table 3 in Box 1). With respect to lifetime fitness consequences, cooperative behaviour must provide direct fitness benefits, i.e. be a mutually beneficial behaviour (table 1 in Box 1), to be under positive selection.

The distinction between behaviour and the underlying strategy

A final important issue for our terminology is that we distinguish between behaviour and underlying strategy.

We use the term 'strategy' loosely, following Maynard Smith (1982) who described a strategy in a most general way: the individual phenotype. Therefore, a strategy is the specification of what an individual will do in any situation in which it may find itself (Maynard Smith, 1982). We do not distinguish between genetically determined strategies and learned strategies (tactics) because of two reasons. First, evolutionary stable strategies (ESS), developmental stable strategies (DSS) and cultural stable strategies (CSS) seem to be conceptually quite similar in their basic forms (Maynard Smith & Price, 1973; Dawkins, 1980; Maynard Smith, 1982). Second, we usually lack information on how genes and learning interact to produce behaviour in a specific situation. Many evolutionary game theoretic models on cooperation explore 'strategies' but assume/allow that behaviour is learned (Axelrod & Hamilton, 1981; Nowak & Sigmund, 1990, 1998).

As long as strategies are unconditional ('always cooperate' and 'always defect'), they are equivalent to the behaviour they produce. However, as soon as individuals are able to flexibly react to changing conditions or when they use information to make decisions about their behaviour, the strategy becomes conditional. For conditional strategies, we have to specify the 'decision rules': what makes an individual decide to show behaviour A instead of behaviour B? For tit-for-tat, the decision rule specifies that a player cooperates if the partner cooperated in the last interaction, and that she/he cheats if the partner cheated in the last interaction (Axelrod & Hamilton, 1981). Cooperative strategies do not necessarily produce cooperative behaviour. For example, a tit-for-tat player matched against a cheater will cooperate in the first round but never thereafter. Also, altruistic behaviour may be conditional if it depends on a recognition mechanism for kin or on greenbeard alleles. We define different strategies in Table 2.

Four fundamental approaches to the evolution of cooperation

Evolutionary pathways

In this section, we describe the answers to the most fundamental functional question about helping behaviour, namely the possible selective forces (to which we apply the term 'evolutionary pathways') that affect the inclusive fitness of the actor in such a way that the helping behaviour is under positive selection.

The evolutionary pathways that allow cooperation and altruism to be selected for are most thoroughly presented in Lehmann & Keller (2006) and West *et al.* (2007). Lehmann & Keller (2006) use mathematical arguments to define key pathways of cooperation and altruism. They are interested in the *average life time fitness consequences* of a behaviour. Helping, like any other behaviour, can only be under positive selection if it increases on average the inclusive fitness of the actor. As the inclusive fitness of an individual consists of the sum of direct and indirect fitness, all models on the evolutionary pathways promoting helping behaviour can be subsumed into two main classes of models. In the first class of models, helping is selected for because it increases the direct fitness of the actor (the gene(s) coding for the behaviour), while in the second class of models, helping is selected for because it increases the indirect fitness of the actor (the gene(s) coding for the behaviour). We refer to direct benefit models as models of cooperative behaviour and to indirect benefit models as models of altruistic behaviour. Models of cooperative behaviour can be subdivided into two classes. An actor may benefit a recipient either (1) because the actor gains a direct fitness benefit from its action, which does not depend on the recipient's response, or (2) because the investment was made in expectation of a future return benefit that provides direct fitness benefits greater than the costs of the initial investment. Likewise, models of altruistic behaviour can be subdivided into two categories.

Table 2 Definitions of various strategies.

Strategy	A strategy is a specification of what an individual will do in any situation in which it may find itself (Maynard Smith, 1982). Strategy will mean the same as behaviour if the strategy produces a fixed behaviour like 'always invest' or 'always cheat'
Decision rule	For conditional strategies, where the behaviour depends on the current state of the individual or its own or the partner's past behaviour, the presence/absence of observers, the decision rule specifies the conditions that will cause an individual to choose a specific behaviour from its available options in a given round
Cooperative strategy	A strategy which, if played against itself, will increase the average payoff (and hence the direct fitness) of the actor and of its partner
Unconditional cooperative strategy	A strategy that will increase the average payoff (and hence the direct fitness) of the partner independently of how this partner behaves
Conditional cooperative strategy	A strategy that causes its bearer to start cooperatively but to respond to a cheating partner in a way that the partner's final payoff will be lower than if it had cooperated
Cheating strategy	A strategy that causes its bearer to maximize its payoff in each current round at the same time diminishing the partner's payoff

Altruistic behaviour prevails if (3) the actor makes an investment that reduces its direct fitness but this cost is more than compensated for by the recipient's gain in indirect fitness benefits through kin selection, or (4) if the actor makes an investment because of a linkage between altruistic behaviour and a phenotypic trait that allows the individual to direct an investment towards others that share the same allele(s) (the green beard scenario; Hamilton, 1964; Dawkins, 1976). Helping behaviour due to green beards is inherently instable because any mutant with the trait but without the helping allele will reap the benefits without incurring the costs and therefore have a selective advantage (Roberts & Sherratt, 2002; Lehmann & Keller, 2006).

The key conclusion to be taken from Lehmann & Keller (2006) and West *et al.* (2007) is that as long as we focus on evolutionary pathways, concepts are reasonably simple: the condition that allows helping to evolve is an increase in the inclusive fitness of the individual or the gene coding for the behaviour, and for this condition to be fulfilled, either the direct or the indirect fitness of the actor (or for the gene coding for the behaviour) must be increased relative to the average fitness in the population. It is important to note that neither modern group selection (for an excellent discussion on the history of multi-level selection see Okasha, 2006) nor network reciprocity (Lieberman *et al.*, 2005; Nowak, 2006) offer alternative pathways to explain helping behaviour. West *et al.* (2007) show that in group selection models based on 'weak altruism' (Wilson, 1980, 1990), individual investors have a higher direct fitness than noninvestors on the population level. Therefore, under such conditions every individual should invest due to self-serving reasons; there is no stable equilibrium that would favour the coexistence of noninvestors. Alternatively, there is only one viable solution to a behaviour that reduces the direct fitness of its actor relative to the population average: there must be compensation due to an increase in the indirect fitness (Lehmann & Keller, 2006). Indirect fitness can only increase due to a sorting mechanism that allows preferential investment towards relatives or individuals that share the gene in question: kin recognition mechanisms, philopatry/limited dispersal or green beard mechanisms. As Lehmann & Keller (2006) point out, trait group selection models and network reciprocity among other papers (table 3 in Lehmann & Keller, 2006) claiming to have found a new pathway to stable cooperation or altruism have actually found a new ecological context (demographic or environmental stochasticity) that permits cooperation or altruism based on individual selection or green beard selection (see next section).

Ecological contexts that facilitate cooperation and altruism

Scientists interested in this aspect of the evolution of helping behaviour explore the ecological conditions that may cause helping to increase the inclusive fitness of the actor. How comes for example that the actor is more related to recipients of its altruistic behaviour than to the average individual in the population? How is it possible that individuals live under conditions that allow for repeated interactions? The helping behaviour itself is usually assumed to be totally unconditional or conditional on external features of the recipient rather than conditional on the behaviour of the recipient.

The major contribution of ecology to cooperation theory concerns the evaluation of conditions that select for the evolution of unconditional helping behaviour. We use the term 'ecology' broadly defined as the environmental conditions in which the individual lives, including conspecifics (i.e. the social environment). Important environmental factors include fluctuations in food or shelter, abundance of predators or the partner or competitor species. Such environmental factors shape to a large extent the social environment such as, e.g. demography, sex ratio, dispersal or group size. The latter can in combination with resource distribution shape competition (scramble vs. contest, within group vs. between groups, within species vs. between species). All the parameters will be reflected in the life history of a species.

In models of altruistic behaviour, indiscriminate helping may be selected for if there is a mechanism that causes a nonrandom distribution of individuals such as limited dispersal, which causes neighbouring individuals to be more related to each other than to the average individual of the population (Hamilton, 1964). Arguably, the most debated ecological context that may favour helping because of direct and/or indirect benefits is group living. Group living animals often experience conflicts between members of the same group but in addition members of the same group may be partners in conflicts with individuals of other groups. The outcome of between group conflicts may have serious consequences for the fitness of the members both of the winning and of the losing group. Access to food sources or refuges from predators or from a harsh environment may be key determinants of an individual's fitness that are gained or lost by all group members, depending on how well they can defend/expand their territory. In primatology, it has long been recognized that severe between group competition may foster increased tolerance or cooperation to resolve within group competition (Wrangham, 1980; van Schaik, 1983; Sterck *et al.*, 1997): a dominant may increase its inclusive fitness by sharing food or reproduction with subordinates if this makes the group more competitive compared to other groups. Benefits of grouping may select for unconditional cooperative behaviour, i.e. in the absence of kin-based benefits of helping. This form of cooperation has been termed 'weak altruism' (Wilson, 1980, 1990) to emphasize that within a group cooperators have a lower fitness than noncooperators. Wilson (2008) maintains the position

that defining altruism as a behaviour that increases group fitness but which reduces the actor's fitness relative to other group members, while defining the selective forces maintaining the behaviour as 'group selection' has let to new insights due to a new perspective. We agree that between-group competition is likely to be a major selective force on the behaviour of social animals, particularly humans (see Boyd *et al.*, 2003; Gintis *et al.*, 2003). Nevertheless, as selection still works on the inclusive fitness of individuals (relative to the average inclusive fitness of individuals in the population) rather than on the total fitness of the group, the term 'group selection' is misleading. Helping simply cannot evolve if it decreases the actor's inclusive fitness (compared to the average population level) even if it increases the fitness of the actor's group. The term 'group competition' captures the conditions that promote unconditional contributions to public goods (that will benefit nonrelatives) in a much clearer way. As long as helpers have a higher direct fitness than nonhelpers, an individual's contribution to a public good is not 'altruistic' but self-serving and hence either what Brown (1983) termed by-product mutualism or a form of pseudo-reciprocity (Connor, 1986): both concepts have in common that an individual gains by investing in group benefits due to the synergistic effects of group living *independently of how other group members behave* (West *et al.*, 2007).

A variety of concepts, each using a different set of terminology, explore how between-group competition leads to cooperative behaviour while avoiding the terms 'altruism' and 'group selection': West *et al.* (2006a) distinguish 'local competition' (within group) from 'global competition' (between groups) and investigate how the relative importance of the two influence cooperative behaviour in humans. Roberts (2005) coined the term 'interdependence between stake-holders' to describe conditions that select for cooperative behaviour. He proposed that the 'r' in Hamilton's famous formula (Hamilton, 1964) can be interpreted as the degree of relatedness between investor and recipient (the conventional interpretation) but also more generally as every interdependence that arises when individuals have an interest or 'stake' in the fitness of the beneficiary, because their own fitness depends on the well-being of the receiver. In cooperative breeding, a mechanism promoting unconditional cooperative behaviour based on interdependence is group augmentation (Kokko *et al.*, 2001).

We will not go into detail discussing the various ecology-based models that allow unconditional helping to be selected for because of unconditional direct benefits, kin selection or green beard selection but instead refer to Lehmann & Keller (2006). While the ecological conditions that promote cooperative or altruistic behaviour are diverse, the mechanisms by which the inclusive fitness of an individual is increased remain the same as in the field interested in evolutionary pathways: direct and/or indirect fitness benefits (Lehmann & Keller, 2006). Also the evolution of 'strong reciprocity' (Fehr & Gächter, 2002) is based on indirect benefits as strong reciprocity depends on low dispersal as ecological condition, causing individuals who punish noncontributing individuals to preferentially inflict costs on nonkin. This form of spite yields indirect benefits because kin get a selective advantage due to 'strong ferocity' (Gardner & West, 2004; Gardner *et al.*, 2007; Lehmann *et al.*, 2007). Spatial reciprocity and trait group selection are other concepts where the stability of unconditional investment is implicitly based on either unconditional direct benefits or limited dispersal leading to kin interactions (Lehmann & Keller, 2006). This may lead to kin selected helping as long as the effects of kin competition do not outweigh the benefits of helping kin (West *et al.*, 2002).

Strategies

Game theory provides the tools for scientists interested in the strategies/decision rules that underlie helping behaviour and that may explain its evolution and stability. Variation of cost or benefits from helping and number of interactions between partners are factors that determine the control mechanisms a player may use to prevent a partner from cheating. The key contributions of game theory include (1) an emphasis on the fact that an individual's best behavioural option may depend on how the partner(s) behave, (2) a framework that may explain variable behaviour both within individuals and between individuals, and (3) the possibility to explore the evolutionary dynamics of cooperative strategies, i.e. the specification of conditions that allow cooperative strategies both to evolve and be maintained. We think this aspect has been neglected in the recent conceptual papers by Lehmann & Keller (2006) and by West *et al.* (2007). Therefore, this approach will be described in much more detail than the other approaches.

Scientists interested in the evolutionary pathways that allow helping to be selected for investigate the average consequences of a behaviour on lifetime direct and indirect fitness, keeping the behaviour of the recipient constant (Lehmann & Keller, 2006; West *et al.*, 2007). However, behaviour is flexible. Therefore, the behaviour of one individual should be dependent on the behaviour of other individuals rather than being fixed. Evolutionary game theory (Maynard Smith & Price, 1973; Maynard Smith, 1982) is the tool to explore the strategies that must necessarily underlie such flexible behaviour and which cause the adjustment of behaviour in response to variable outcome of interactions. Trivers (1971) provided the starting point for all research that seeks to understand how cooperative behaviour can be enforced with his concept of reciprocal investment ('reciprocal altruism'). The basic observation is that we can often observe behaviours with the *immediate* effects of a benefit for the recipient and a cost to the actor. For example, playing 'C'

in any round of an iterated prisoner's dilemma is an investment: by definition the payoff within the interaction is lower than if the player had defected (Luce & Raiffa, 1957). Only the partner's future behaviour may more than compensate for this investment. Therefore, the question emerges how the investor ensures that it will gain future benefits from the investment that will more than compensate the current costs as otherwise this form of helping would be counter-selected.

Here, we focus on individual strategies that promote cooperation as there is hardly any research on altruistic strategies. Cooperation often occurs between related individuals (Clutton-Brock, 2002; West *et al.*, 2002) but the relevant models usually assume that partners are unrelated to each other or that indirect benefits are not high enough to promote helping behaviour. We will not distinguish between intraspecific cooperation and interspecific mutualism, neither with respect to the models nor with respect to examples that we will provide as illustrations, as the general problem remains the same: we want to know how investments may provide more than compensatory benefits to the actor. More specifically, we want to know how the strategies of investors ensure return benefits with cooperative partners or reduce both own losses and the gains of a cheating partner (Bshary & Bronstein, 2004; Sachs *et al.*, 2004; Noë, 2006). We therefore have to ask (1) how the investment affects the behaviour of the recipient or of bystanders and (2) how investors behave in a similar situation in the future, depending on what their original investment has yielded.

Two main approaches have been used in evolutionary game theory. One approach is to first specify the game structure, then to think about possible strategies and finally to let the strategies compete against each other in computer simulations to determine whether one or more cooperative strategies could be evolutionarily stable (Axelrod & Hamilton, 1981). The second approach is to identify control mechanisms that may favour cooperative partners while inflicting costs on cheating partners, and to explore the range of conditions in which the control mechanism may yield stable cooperation. We will now present these two approaches in more detail.

Searching evolutionarily stable cooperative strategies

The classic study for this kind of approach by Axelrod & Hamilton (1981) provided the possibility to analyse the evolutionary stability of competing strategies in the iterated prisoner's dilemma game. They first specified the game structure: players are paired randomly, each player has two behavioural options, a payoff matrix specifies the payoff for each player in each possible combination of behaviours, and there is a fixed probability of playing another round with the same partner. In the second step, they asked colleagues to submit strategies that they thought to be competitive in the specified game structure. Finally, they ran a computer tournament where the average payoff of a strategy in one round of interactions translated into the strategy's abundance (increasing or decreasing in relative frequency) in the next round. Axelrod & Hamilton (1981) found two 'winners', either 'always defect' or 'tit-for-tat', a simple conditional cooperative strategy that causes the individual to start cooperatively in the first round and then to copy the partner's behaviour of each previous round. Thus, a tit-for-tat player cooperates as long as the partner cooperates but switches to cheating if the partner cheats. Meanwhile, new conditional cooperative strategies have been tested in the iterated prisoner's dilemma and emerged as superior to tit-for-tat (references in Dugatkin, 1997).

Maynard Smith (1982) noted that a weakness of the research using game theory at that time was that too much emphasis was made on finding stable equilibria rather than trying to define the phenotype set, i.e. the strategies that could be used by players. The variation in potential strategies may be limited by constraints on physiology, lack of information or lack of cognitive abilities, among others. To give some examples, a noncompetitive individual cannot reasonably threaten to inflict harm on a noncooperative dominant; selectively helping cooperative individuals requires close spatial association so that individuals could in principle acquire the necessary information, and also requires strong memory capacities. It is therefore a key challenge for both theoreticians and empiricists to identify all potential strategies that could be played in specific case studies, applying their knowledge about the system to identify constraints and elucidate the game structure itself. Unfortunately, this effort has rarely been made. Only for the iterated prisoner's dilemma game, a large variety of strategies has been developed and tested against each other (Axelrod & Hamilton, 1981; Boyd, 1989; Nowak & Sigmund, 1992, 1993). However, constraints on playing any one of these strategies have rarely been addressed (but see Milinski & Wedekind, 1998) and the game structure seems to apply to very few known examples of cooperation (Dugatkin, 1997), while they apparently are irrelevant for the many known cases of mutualisms (Bergstrom *et al.*, 2003). For other games, few strategies have been tested and potential constraints have been ignored. For example, we are not aware of any study where various strategies based on punishment as control mechanism must compete against each other to see which one prevails. Just to name a few possibilities, a punishment strategy could be 'always cooperate and punish your partner after each round in which it failed to cooperate as well', or 'play tit-for-tat and in addition punish the partner for each defection' or 'start cooperatively, punish your partner the first time it fails to cooperate and switch to defection if the punishment does not alter the partner's behaviour'. To identify feasible

strategies and their respective potential constraints remains a key challenge for games other than the prisoner's dilemma.

The controlling components of cooperative strategies

A crucial component of a strategy is how individuals foster cooperative behaviour on the part of the interaction partners or how they prevent cheating of the partner. Individuals may for example match the partner's current behaviour in the next interaction (like tit-for-tat does) or they could respond to cheating with aggression or with the termination of the relationship. Such reactions are called control mechanisms because they will have negative effects on the total payoff of cheaters. Analytical models can be used to specify conditions under which a particular control mechanism may yield stable cooperation. Over the last 20 years or so, a large variety of concepts that may explain stable cooperation have been developed. In the literature, one can find by-product mutualism, pseudo-reciprocity, group augmentation, pay-to-stay, reciprocity, threat of reciprocity, parcelling, punishment, sanctions, power, partner switching, generalized reciprocity, strong reciprocity, policing, indirect reciprocity and social prestige. This diversity results partly because some terms are synonymous. However, it has also become clear from empirical advances that we need a large variety of concepts to grasp all the known examples of cooperation and mutualism (Bergmüller *et al.*, 2007a). In an attempt to point out similarities and differences between each concept, Bergmüller *et al.* (2007a) found that most of the concepts can be classified with a combination of four basic parameters where each can be in one of two different states. For a detailed discussion of this classification we refer to the original paper as well as to 22 commentaries and the authors' reply (Bergmüller *et al.*, 2007b) in a special edition of Behavioural Processes (2007). Below, we restrict ourselves to a brief overview.

The following four parameters can be seen as building blocks to define the controlling aspect of a strategy that ensures that (within a certain parameter space) helping yields on average a net fitness benefit for the helper. (1) The act of helping: an investment or a self serving mutually beneficial behaviour? (2) The return benefits: an investment (i.e. a costly response) or a self serving mutually beneficial behaviour? Reciprocity is defined by mutual investment, whereas in pseudo-reciprocity the behaviour of one player is self-serving mutually beneficial. (3) Identity of the individual that provides the return benefits: the recipient or a bystander in a communication network (McGregor, 1993)? We call the former a 'direct response' and the latter an 'indirect response' (following Nowak & Sigmund, 1998). From the perspective of the responding individual, it might be more useful to describe direct benefits as experience based and indirect benefits as information based (Roberts & Sherratt, 2007). Note that this use of 'direct response' and 'indirect response' should not be confused with the 'direct benefits' and 'indirect benefits' through which the inclusive fitness of the actor is increased. (4) The nature of the return benefits: due to receiving a reward or due to avoiding a cost? Following Clutton-Brock (2002) we use the adjective 'positive' for the former and 'negative' if failure to help causes the infliction of a cost ('punishment'). A combination of the states of the four parameters yields nine different basic concepts (Fig. 1, adapted from Bergmüller *et al.*, 2007a).

The nine basic concepts that may explain why helping leads to direct fitness benefits for the actor

1. *By-product beneficial behaviour*. In this simplest form of cooperation, the mere existence of other individuals and their self-serving actions provide benefits to others, without involving investments. Its evolution and stability is therefore straightforward (Dugatkin, 1997; Leimar & Connor, 2003). Examples include cooperative hunting in jackals (Lamprecht, 1978) and more generally apply to cases of coordination (Clutton-Brock, 2002). Coordination is the basis for group living (selfish herd, Hamilton, 1971), mixed species associations and interspecific coordinated hunting (Bshary *et al.*, 2006). Also some cases of group augmentation (Kokko *et al.*, 2001) such as self-serving contributions to public goods (West *et al.*, 2007, in their re-evaluation of 'weak altruism') fulfil the criteria of cooperation.

2. *Direct positive pseudo-reciprocity* (='pseudo-reciprocity', 'group augmentation'). In pseudo-reciprocity the recipient will use an investment for its own benefits. The

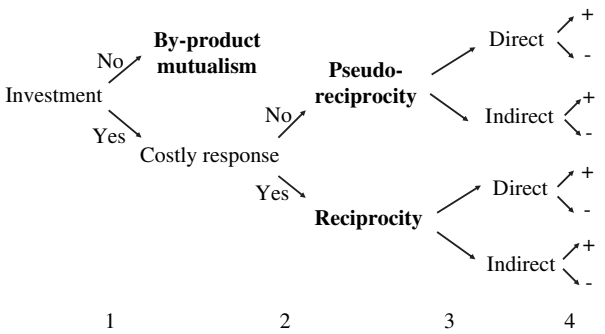

Fig. 1 Hierarchical classification of mechanisms that can maintain cooperative behaviour. By-product mutualism does not involve (1) investments that are directed towards others. An investment may be performed to obtain benefits resulting from the self-serving behaviour of the receiver (i.e. pseudo-reciprocity), without eliciting return investment. Alternatively, an investment may be (2) made in expectation of an investment in return (costly response), resulting in reciprocity. The investor may obtain benefits (3) either directly or indirectly (i.e. via third parties). (4) Cooperative behaviour may be stabilized by costly acts or by-products resulting from self-serving responses by the receiver (or third parties) that have either positive (+) or negative (−) effects on the partner.

donor benefits because the self-serving behaviour of the receiver benefits the investor as a by-product (Connor, 1986). The concept of group augmentation includes the very same logic. Most ant-mutualisms appear to be cases of pseudo-reciprocity (Leimar & Connor, 2003): the partner species typically invests in providing food rewards, which causes the ants to self-servingly defend their food sources against their predators.

3. *Direct negative pseudo-reciprocity*. This control mechanism relies on the potential victim's ability to terminate the interaction (self-servingly), which has negative effects for the potential exploiter. The two basic concepts are 'power' (Johnstone & Bshary, 2002; Bowles & Hammerstein, 2003) and 'sanctions' (Herre *et al.*, 1999; Kiers *et al.*, 2003). Sanctions work because of a *sequential* game structure. One class of players makes an initial investment that is available to members of another class of interaction partners that have to make their offer. The initial investor can then selectively stop the interaction if a partner did not offer net benefits, through which the partner loses everything. Experimental evidence for sanctions has been provided in leguminose plant–rhizobia interactions, where plants selectively stop the maintenance of nodules in which the bacteria fail to fix a minimum amount of nitrogen (Kiers *et al.*, 2003). Power differs from sanctions in that actions are not sequential but *parallel*. Many real life interactions usually last some time, allowing a potential exploitee to prematurely end an interaction. This selects for potential cheaters to cooperate as long as the payoff of a prolonged cooperative interaction is higher than the payoff of a shorter exploitative interaction. Both power and sanctions may yield cooperative outcomes in one-off interactions.

A third form of negative pseudo-reciprocity is partner switching. If a player cheats, the victim's best option may be to switch to another partner for the next interaction (Ferrière *et al.*, 2002; Bshary & Grutter, 2002a). Partner switching requires a repeated game structure and an asymmetry between cheater and victim: the cheater must belong to the abundant class of players from which individuals are chosen by potential victims, which are the members of the rare class of players. Under these circumstances, leaving a cheater is self serving while the cheater incurs a cost because it will spend some time without any interaction partner. Client reef fish with access to several cleaning stations appear to use switching as a mechanism to control the behaviour of cleaner wrasses (Bshary & Schäffer, 2002).

4. *Indirect positive pseudo-reciprocity*. This concept is based on 'social prestige' (Zahavi, 1995; Roberts, 1998; Lotem *et al.*, 2003). In social prestige individuals signal their quality (cooperative behaviour is a handicap) to bystanders through helping. Bystanders choosing to interact with individuals with high prestige make a self-serving decision; they can expect personal benefits from this choice, like females choosing a high quality male to sire her offspring. In cleaning mutualism involving the cleaner wrasse *Labroides dimidiatus*, clients pay attention to how cleaners treat their current client and cleaners are therefore more cooperative towards their current client in the presence of bystanders (Bshary & Grutter, 2006). Clients are self-serving in choosing to interact with a cleaner that treated another client well and avoiding interactions with a cleaner that cheated another client as they make their choice in order to increase the average service quality they receive.

5. *Indirect negative pseudo-reciprocity.* The concept applies to situations where an actor helps a recipient because otherwise a third party individual would do best by evicting the actor from the area. The concept could be applied to helpers that 'pay-to-stay' (Gaston, 1978) in cooperative breeding: helpers invest in offspring because otherwise it would be in the self interest of the breeder to evict the helper. However, it is a matter of perspective whether the helper actually helps the offspring or the breeder to avoid eviction from the territory. In the latter case, the helper would provide food to avoid direct negative pseudo-reciprocity (as it has been classified by Bergmüller *et al.* (2007a), but see Gilchrist (2007).

6. *Positive direct reciprocity* (= 'reciprocity', reciprocal altruism', 'reciprocal investment', 'parcelling'). The controlling aspect of positive reciprocity is based on rewarding cooperative partners: as long as the partner invests, the focal individual invests in return. If the partner cheats, however, the focal individual switches to cheating as well in the next round. Tit-for-tat and its cousins (Dugatkin, 1997) are the key strategies for repeated game structures. A special case is 'parcelling' (Connor, 1986, 1995), where partners cut the total investment into pieces and transfer a shot prisoner's dilemma into an iterated game. The classic example is the egg trading in hamlet fish, a simultaneous hermaphrodite (Fischer, 1988).

7. *Negative direct reciprocity* (='punishment'). This control mechanism is based on an individual inflicting costs on a noncooperating partner at own expenses. Punishment therefore reduces the immediate payoff of the punisher (Clutton-Brock & Parker, 1995). In contrast to sanctions, punishment can therefore only evolve in a repeated game structure (unless it provides indirect fitness benefits, see Gardner & West, 2004). The function of the act is to alter the future behaviour of the victim towards cooperative behaviour, which will then benefit the punisher. An empirical example based on experimental evidence are client reef fish that respond to cheating by cleaners with aggression, which causes cleaners to behave more cooperatively towards the same client in their next interaction (Bshary & Grutter, 2002b, 2005). Also the pay-to-stay concept may be a form of negative direct reciprocity if the breeder punishes a noncontributing helper rather than evicting it. Only future empirical studies can reveal the relative importance of

punishment and eviction for stable contributions of helpers in cooperatively breeding species where helping is based on pay-to-stay.

8. *Indirect positive reciprocity* (='image scoring', 'generalized reciprocity'). In indirect reciprocity based on image scoring, individuals invest only in partners that have sufficiently helped others in the past (Alexander, 1987). Helping raises the 'image score' while failure to help reduces the score. An image score above a critical threshold is necessary to receive help from third parties (Nowak & Sigmund, 1998; Leimar & Hammerstein, 2001). Empirical evidence for indirect positive reciprocity based on image scoring is currently restricted to humans (Wedekind & Milinski, 2000).

Another form of indirect positive reciprocity is generalized reciprocity. In this game, the logical order of reasoning is reversed: rather than investing in order to receive benefits in the future, individuals that received help are willing to invest into third parties. The identity of the third party or the third party's past behaviour do not influence decisions; players only need to know what happened to themselves rather than how potential recipients behaved in the past (Pfeiffer *et al.*, 2004; Hamilton & Taborsky, 2005). First evidence for this concept has been provided in rats (Rutte & Taborsky, 2007).

9. *Indirect negative reciprocity* ('policing', 'strong reciprocity'). Indirect negative reciprocity is also called policing. Policing occurs in hymenoptera where workers eat the eggs laid by other workers and attack these 'cheaters' (Ratnieks & Wenseleers, 2005). However, it is unclear how relatedness between individuals influences policing, therefore kin selection might be involved. Indirect negative reciprocity has also attracted much attention in studies on human behaviour (Fehr & Gächter, 2002), as humans are willing to pay money in order to punish individuals who behaved uncooperatively towards others in one-shot games under anonymous laboratory conditions ('strong reciprocity').

The social scientists' approach

Game theoretic analyses may predict which conditional strategy should be used to ensure revenues under specific circumstances. However, it might not be possible for the individuals involved to play a certain strategy because of cognitive limitations. For instance, tit-for-tat players may need to individually recognize partners and keep track of past interactions with these partners (book keeping), which requires some learning and memory capacities (Hammerstein, 2003b). Social evolutionary scientists have introduced a research field to cooperation and altruism that focuses on decision making in humans. Differences in cognitive abilities are often used as argument why humans should be able to cooperate with unrelated individuals on a scale that is apparently unmatched in other animals (Fehr & Gächter, 2002; Gintis *et al.*, 2003). It is proposed that human cooperation is often special with respect to the complexity of mechanisms that govern human decisions such as emotion, language, 'high level' culture, norms, moral judgement or long term memory (Fehr & Fischbacher, 2003; Gintis, 2006).

It is important to note that differences in cognitive abilities are differences with respect to the mechanisms underlying behaviour, which do not lead to differences with respect to the evolutionary pathways that promote cooperation (West *et al.*, 2007). Nevertheless, the fact that many humans are willing to reward cooperative third parties and to punish egoistic third parties in anonymous one-shot interactions (Fehr & Gächter, 2002; Fehr & Fischbacher, 2003) is certainly intriguing. Such 'strong reciprocators' leave the experiment with less money in their pocket than individuals who cooperate in direct interactions but do not reward or punish third parties. While the propensity to show such behaviours must yield either direct or indirect benefits under natural conditions (Gardner & West, 2004; Lehmann *et al.*, 2007), we clearly need to understand how such behaviours can be promoted on the proximate level. Scientists working on animals should join the social scientists' quest of understanding decision making processes. Species in which behaviour is primarily genetically determined may easily evolve an appropriate strategy (West *et al.*, 2006b). But in species where learning plays a role, the type of behaviour that is learned is typically influenced by short-term reward and punishment, i.e. through operant conditioning (Wynne, 2001), which leads to strong discounting of the future (Wynne, 2001; Stephens *et al.*, 2002). Therefore, as long as individuals receive a larger short-term benefit if they do not invest or punish, it is not immediately evident how they should learn to behave as a strong reciprocators because of delayed benefits. Humans apparently discount the future much less than other animals tested so far (Wynne, 2001), which means that their cognitive constraint on accepting low immediate payoffs in favour of high delayed payoffs is much smaller than that of other animals. In addition, humans have evolved a powerful alternative to immediate *material* benefits: a highly evolved neocortex provides humans with emotions and the ability to develop moral judgement, which in turn provides humans with self rewarding immediate *psychological* benefits for helping and for punishment of transgressors (de Quervain *et al.*, 2004). Other proximate mechanisms of cooperation in humans seem to be deeply rooted and beyond conscious assessment. For example, humans behave more cooperatively in the presence of artificial eyes (Haley & Fessler, 2005; Bateson *et al.*, 2006). Also hormones can mediate human cooperative behaviour. For example, oxytocin increases trust in partners (Kosfeld *et al.*, 2005). Such mechanisms can be expected to play a role across the animal kingdom. However, only future studies may reveal their importance and limitations.

Discussion

We attempted to clarify four major approaches to understand the evolution and persistence of cooperation and altruism: evolutionary pathways, ecological settings, conditional strategies, and the social scientists' approach. Similar to the well-known '4 why-questions' in biology (Tinbergen, 1963), these questions should be seen as complementary, not as alternatives. We showed that the four approaches differ with respect to the 'conditions' and the 'mechanisms' they study. As a consequence, one may study helping with regards to the lifetime or the immediate fitness consequences for the actor. In addition, depending on the approach taken one may be interested primarily in the social behaviour of individuals or additionally in the outcome of interactions. We believe that by highlighting these distinctions the different perspectives become apparent, so that we can reduce potential misunderstandings among researchers working on different aspects and perspectives of helping. Only considering all approaches will allow us to obtain a comprehensive understanding of cooperation and altruism and the potential differences between humans and other species. Besides, we should not forget that the other 'why' questions of Tinbergen (1963), namely questions about the phylogeny and the ontogeny of cooperative and altruistic behaviour provide additional important insights.

Some examples of how the four approaches can be confused

We believe that a key source for confusion in the cooperation/altruism literature is that discussions mix evolutionary pathways with ecological contexts, individual strategies and cognitive mechanisms. Lehmann & Keller (2006) shed light on the existing confusion between basic social evolution theory and the ecological approach. West *et al.* (2007) further clarify this issue and explain in detail how the social scientists' approach can be confused with basic social evolution theory. Here, we want to highlight in particular the confusion over strategies. In the only textbook on animal cooperation, Dugatkin (1997) tries to classify all known examples according to four categories: by-product mutualism, reciprocity, kin selection and group selection. These four categories are a mixture of two strategies (by-product mutualism and reciprocity), one evolutionary pathway (kin selection), and one ecological condition (between group competition). We illustrate the confusion with two examples, namely kin selection and reciprocity. First, kin selection describes an evolutionary pathway that explains altruistic behaviour. The ecological conditions for kin selection most generally involve a mechanism for nonrandom assortment of individuals such as limited migration so that individuals interact in part with relatives. On the strategic level, helping could be conditional on relatedness or indiscriminative. On the cognitive level, conditional helping requires some rules how an individual may recognize relatives, such as growing up together or having a similar smell. Reciprocity is a strategy while the evolutionary pathway is 'enforced direct benefits' according to West *et al.* (2007). Ecological conditions may favour longevity and reduced movement, which in turn favours (positive direct) reciprocity because this strategy requires a repeated game structure to be evolutionarily stable. Finally, reciprocity requires some cognitive abilities as individuals must remember certain aspects of past interactions.

Similarly to Dugatkin, Nowak (2006) distinguishes five major 'mechanisms' for the evolution of cooperation: kin selection, direct reciprocity, indirect reciprocity, network reciprocity and group selection. Direct reciprocity and indirect reciprocity describe the controlling aspect of strategies that make sure (under certain conditions) that an investment yields on average higher returns. Kin selection is one main path to promote helping (here: altruistic) behaviour but several strategies could yield the benefits. Ecological circumstances that lead to between group competition or limited migration may in turn promote direct benefits or kin selection. Both 'group selection' and network reciprocity explore such ecological parameters (Lehmann & Keller, 2006).

Finally, West *et al.* (2007) (1) show that cognitive mechanisms should not be confused with evolutionary pathways, and (2) intend to counter the general perception that kin selection and reciprocity are the two main concepts to explain helping behaviour. We fully agree as kin selection and reciprocity are not complementary concepts but they address helping on two different levels (pathways and strategies). To match levels, one could either propose that direct enforced benefits and kin selection are the two main pathways to explain helping, or one could propose that reciprocity and helping individuals that are raised by one's mother (or alternatively who grew up on the same territory or who smell similarly) are the two main strategies to explain helping. We would certainly disagree with the latter statement as evidence for direct reciprocity is scarce (Dugatkin, 1997; Bergstrom *et al.*, 2003).

Links between the four research fields

While we argue that it is important to distinguish between the four fields, it is clear that the four fields are partly linked and that future progress depends on further integration. The most basic common denominator is that all concepts are linked to basic social evolution theory, as they ultimately explore the potential of direct and/or indirect fitness benefits to cause selection on

helping. The ecological approach is most tightly linked to basic evolution theory, as becomes apparent from a recent target article with responses (Lehmann & Keller, 2006).

Ecology and game theory are implicitly linked as the ecology of a species translates into the game structure (Bshary & Bronstein, 2004). The number of rounds two individuals may on average play with each other depend on longevity and migration patterns. A payoff matrix is ideally a reflection of the ecological conditions. Game theoreticians still have to build their models more on the empirical evidence in order to make the models more applicable/testable. For example, we need models where the payoff can be variable, as it happens in nature when ecological conditions change. It is intuitively obvious that exact payoffs depend on the condition of the player(s) involved, i.e. hunger level or physical strength. In interspecific mutualisms, the service provided by one partner often depends on the population dynamics of a third species. For example, ants provide their various partner species with protection against predators in return for food and/or shelter (Pierce *et al.*, 2002; Heil & McKey, 2003). The partners attract ants with food and hence have rather fixed expenses. The ants, however, can only provide a service if their partner is actually attacked by a predator. Therefore in years with a high predator density, ants more than compensate their partners' investments while in years with low predator density they simply cannot. Similarly, cleaner fish may reduce the parasite load of their clients only to an effective degree if the ectoparasite populations are high. In contrast, if ectoparasite densities are low, cleaners might feed more on mucus and hence become increasingly parasitic (Grutter, 1997).

Finally, there are also important links between game theory and the social scientists' approach (the social scientists combined them of course from the beginning). For any game structure, it is important to find out (1) which strategies/partner control mechanisms may promote cooperative behaviour, and (2) what cognitive requirements are necessary to use a strategy/a control mechanism successfully under the specified conditions. For example, Hammerstein (2003b) proposes that reciprocity is rare in nature, but not necessarily because the game structure is rare, but because most animals lack the cognitive requirements of individual recognition and particularly book keeping. However, such potential constraints have rarely been investigated. It remains a vastly open research field to study the decision making processes that cause cooperating, punishing and cheating in animals.

Towards more realistic concepts of helping

A key remaining challenge is to try to understand variation between individuals. Why are some individuals 'unconditional cooperators' and others 'egoists'? In humans one may distinguish different individual types (Ostrom *et al.*, 1999) such as 'free riders', 'cautious cooperators' (only cooperate when return is protected against free riders), 'hopeful cooperators' (initiate investment), and 'unconditional cooperators' (believe in common goods). Such strategies may be inflexible and thus describe a certain type of individual coping style or personality (Wilson *et al.*, 1994; Gosling & John, 1999). Differences between individuals also have been demonstrated in nonhuman animals and are termed 'animal personalities' or 'behavioural syndromes' (Drent *et al.*, 2003; Sih *et al.*, 2004). Lions provide a famous example in the context of cooperative territory defence, where one can distinguish between unconditional cooperators, conditional cooperators, conditional laggards and unconditional laggards (Heinsohn & Packer, 1995).

Consistent individual differences may persist because of constraints (Sih *et al.*, 2004). Sherratt & Roberts (2002) introduced 'phenotypic defectors' in their model, which are individuals that are too weak to be able to help others, and found that the variation in condition stabilizes the persistence of cooperative behaviour (see also Lotem *et al.*, 2003). Alternatively, consistent individual differences may be adaptive in that they are equivalent to certain strategies or niche options (i.e. alternative strategies) within a species or population (Bergmüller & Taborsky, 2007). Although much of cooperation theory implicitly assumes the existence of different explicit and fixed types, such as 'cooperators' and 'defectors', only with the recent increase in focus on the behaviour of individuals has the significance of individual strategies with regards to cooperative interactions started to become elucidated (Arnold *et al.*, 2005; Komdeur, 2006, 2007; Bergmüller & Taborsky, 2007). In this context, a model by McNamara *et al.* (2004) is of major importance. The authors analysed a game where players could potentially interact with the same partner for 100 rounds but the interaction would be terminated if one of them cheated. The ESS solution is a distribution of strategies coding for playing a different number of rounds cooperatively before cheating (McNamara *et al.*, 2004). We clearly need empirical and theoretical studies to establish to what extent behavioural syndromes in animals are linked to different levels of cooperative behaviour.

From what we developed above, we argue that a key shortcoming of most current modelling efforts on cooperation is that the models still focus on symmetric game structures where each individual chooses between the same behavioural options. In Nature, however, many interactions are based on asymmetrical strategy sets (Bshary & Grutter, 2002a; Bshary & Bronstein, 2004; Bergmüller *et al.*, 2007a). For example, a dominant is much more likely than a subordinate to use punishment to enforce cooperative behaviour. In many cases, only

one player has the option to cheat while the partner simply lacks the option (i.e. cannot perform such behaviour). Only a stronger collaboration between theoreticians and empiricists will allow us to ensure that future models are based on real-life examples, yielding testable predictions that provide feedback that may be used to readjust theory.

Acknowledgments

We thank Laurent Keller and Laurent Lehmann for discussion, Laurent Keller, Jenny Oates, Andy Gardner, and an anonymous referee for comments on earlier versions of the manuscript, Jenny Oates for improving the English and Andrea Hohner for help in editing. The study was funded by a grant from the Swiss National Science Foundation.

References

Alexander, R.D. 1987. *The Biology of Moral Systems*. Aldine de Gruyter, New York.

Arnold, K.E., Owens, I.P.F. & Goldizen, A.W. 2005. Division of labour within cooperatively breeding groups. *Behaviour* **142**: 1577–1590.

Axelrod, R. & Hamilton, W.D. 1981. The evolution of cooperation. *Science* **211**: 1390–1396.

Bateson, M., Nettle, D. & Roberts, G. 2006. Cues of being watched enhance cooperation in a real-world setting. *Biol. Lett.* **2**: 412–414.

Begon, M., Harper, J.L. & Townsend, C.R. 2005. *Ecology*. Blackwell, Boston, Oxford.

Bergmüller, R. & Taborsky, M. 2007. Adaptive behavioural syndromes due to strategic niche specialization. *BMC Ecol.* **7**: 72.

Bergmüller, R., Johnstone, R.A., Russell, A.F. & Bshary, R. 2007a. Integrating cooperative breeding into theoretical concepts of cooperation. *Behav. Process.* **76**: 61–72.

Bergmüller, R., Russell, A.F., Johnstone, R.A. & Bshary, R. 2007b. On the further integration of cooperative breeding and cooperation theory. *Behav. Process.* **76**: 170–181.

Bergstrom, C.T., Bronstein, J.L., Bshary, R., Connor, R.C., Daly, M., Frank, S.A., Gintis, H., Keller, L., Leimar, O., Noë, R. & Queller, D.C. 2003. Group report: interspecific mutualism – puzzles and predictions. *Genetic and Cultural Evolution of Cooperation* (Hammerstein, P., ed.), pp. 241–256. MIT Press, Cambridge, MA.

Bowles, S. & Hammerstein, P. 2003. Does market theory apply to biology? In: *Genetic and Cultural Evolution of Cooperation* (Hammerstein, P., ed.), 163–165. MIT Press, Cambridge, MA.

Boyd, R. 1989. Mistakes allow evolutionary stability in the repeated prisoners-dilemma game. *J. Theor. Biol.* **136**: 47–56.

Boyd, R., Gintis, H., Bowles, S. & Richerson, P.J. 2003. The evolution of altruistic punishment. *Proc. Natl. Acad. Sci.* **100**: 3531–3535.

Bronstein, J.L. 2001. The exploitation of mutualisms. *Ecol. Lett.* **4**: 277–287.

Brown, J.L. 1983. Cooperation: a biologist's dilemma. In: *Advances in the Study of Behavior* (J.S. Rosenblatt, ed), pp. 1–37. Academic Press, New York.

Bshary, R. & Bronstein, J.L. 2004. Game structures in mutualistic interactions: what can the evidence tell us about the kind of models we need? *Adv. Study Behav.* **34**: 59–101.

Bshary, R. & Grutter, A.S. 2002a. Asymmetric cheating opportunities and partner control in a cleaner fish mutualism. *Anim. Behav.* **63**: 547–555.

Bshary, R. & Grutter, A.S. 2002b. Experimental evidence that partner choice is a driving force in the payoff distribution among cooperators or mutualists: the cleaner fish case. *Ecol. Lett.* **5**: 130–136.

Bshary, R. & Grutter, A.S. 2005. Punishment and partner switching cause cooperative behaviour in a cleaning mutualism. *Biol. Lett.* **1**: 396–399.

Bshary, R. & Grutter, A.S. 2006. Image scoring and cooperation in a cleaner fish mutualism. *Nature* **441**: 975–978.

Bshary, R. & Schäffer, D. 2002. Choosy reef fish select cleaner fish that provide high-quality service. *Anim. Behav.* **63**: 557–564.

Bshary, R., Hohner, A., Ait–El-Djoudi, K. & Fricke, H. 2006. Interspecific communicative and coordinated hunting between groupers and giant moray eels in the Red Sea. *PLoS Biol.* **4**: 2393–2398.

Cant, M.A. & Johnstone, R.A. 2006. Self-serving punishment and the evolution of cooperation. *J. Evol. Biol.* **19**: 1383–1385.

Clutton-Brock, T. 2002. Breeding together: kin selection and mutualism in cooperative vertebrates. *Science* **296**: 69–72.

Clutton-Brock, T.H. & Parker, G.A. 1995. Punishment in animal societies. *Nature* **373**: 209–216.

Connor, R.C. 1986. Pseudo-reciprocity – investing in mutualism. *Anim. Behav.* **34**: 1562–1566.

Connor, R.C. 1995. Altruism among nonrelatives – alternatives to the prisoners-dilemma. *Trends Ecol. Evol.* **10**: 84–86.

Dawkins, R. 1980. Good strategy or evolutionarily stable strategy? In: *Sociobiology: Beyond Nature/Nurture* (Barlow, G.W. & Silverberg, J., eds), pp. 331–367. Westview Press, Boulder.

Drent, P.J., Van Oers, K. & Van Noordwijk, A.J. 2003. Realized heritability of personalities in the great tit. *Proc. R. Soc. Lond. B Biol. Sci.* **270**: 45–51.

Dugatkin, L.A. 1997. *Cooperation among Animals. An Evolutionary Perspective*. Oxford University Press, Oxford.

Fehr, E. & Fischbacher, U. 2003. The nature of human altruism. *Nature* **425**: 785–791.

Fehr, E. & Gächter, S. 2002. Altruistic punishment in humans. *Nature* **415**: 137–140.

Ferrière, R., Bronstein, J.L., Rinaldi, S., Law, R. & Gauduchon, M. 2002. Cheating and the evolutionary stability of mutualisms. *Proc. R. Soc. Lond. B Biol. Sci.* **269**: 773–780.

Fischer, E.A. 1988. Simultaneous hermaphroditism, tit-for-tat, and the evolutionary stability of social-systems. *Ethol. Sociobiol.* **9**: 119–136.

Gardner, A. & West, S.A. 2004. Cooperation and punishment, especially in humans. *Am. Nat.* **164**: 753–764.

Gardner, A., West, S.A. & Barton, N.H. 2007. The relation between multilocus population genetics and social evolution theory. *Am. Nat.* **169**: 207–226.

Gaston, A.J. 1978. The evolution of group territorial behavior and cooperative breeding. *Am. Nat.* **112**: 1091–1100.

Gilchrist, J.S. 2007. Cooperative behaviour in cooperative breeders: costs, benefits, and communal breeding. *Behav. Process.* **76**: 100–105.

Gintis, H. 2006. A framework for the unification the behavioral sciences. *Behav. Brain Sci.* **30**: 1–16.

Gintis, H., Bowles, S., Boyd, R. & Fehr, E. 2003. Explaining altruistic behavior in humans. *Evol. Hum. Behav.* **24**: 153–172.

Gosling, S.D. & John, O.P. 1999. Personality dimensions in nonhuman animals: a cross-species review. *Curr. Dir. Psychol. Sci.* **8**: 69–75.

Grutter, A.S. 1997. Spatiotemporal variation and feeding selectivity in the diet of the cleaner fish *Labroides dimidiatus*. *Copeia* **2**: 346–355.

Haley, K.J. & Fessler, D.M.T. 2005. Nobody's watching? Subtle cues affect generosity in an anonymous economic game *Evol. Hum. Behav.* **26**: 245–256.

Hamilton, W.D. 1964. The genetical evolution of social behaviour I u. II.. *J. Theor. Biol.* **7**: 1–52.

Hamilton, W.D. 1970. Selfish and spiteful behavior in an evolutionary model. *Nature* **228**: 1218–1220.

Hamilton, W.D. 1971. Geometry of the selfish herd. *J. Theor. Biol.* **31**: 295–311.

Hamilton, I.M. & Taborsky, M. 2005. Contingent movement and cooperation evolve under generalized reciprocity. *Proc. R. Soc. Lond. B Biol. Sci.* **272**: 2259–2267.

Hammerstein, P. 2003a. Genetic and cultural evolution of cooperation. In: *Dahlem Workshop Reports*. The MIT Press, Cambridge.

Hammerstein, P. 2003b. Why is reciprocity so rare in social animals? A protestant appeal. In: *Genetic and Cultural Evolution of Cooperation* (P. Hammerstein, ed), pp. 83–93. MIT Press, Cambridge.

Heil, M. & McKey, D. 2003. Protective ant-plant interactions as model systems in ecological and evolutionary research. *Annu. Rev. Ecol. Evol. Syst.* **34**: 425–453.

Heinsohn, R. & Packer, C. 1995. Complex cooperative strategies in group-territorial African lions. *Science* **269**: 1260–1262.

Herre, E.A., Knowlton, N., Mueller, U.G. & Rehner, S.A. 1999. The evolution of mutualisms: exploring the paths between conflict and cooperation. *Trends Ecol. Evol.* **14**: 49–53.

Johnstone, R.A. & Bshary, R. 2002. From parasitism to mutualism: partner control in asymmetric interactions. *Ecol. Lett.* **5**: 634–639.

Kiers, E.T., Rousseau, R.A., West, S.A. & Denison, R.F. 2003. Host sanctions and the legume-rhizobium mutualism. *Nature* **425**: 78–81.

Kokko, H., Johnstone, R.A. & Clutton-Brock, T.H. 2001. The evolution of cooperative breeding through group augmentation. *Proc. R. Soc. Lond. B Biol. Sci.* **268**: 187–196.

Komdeur, J. 2006. Variation in individual investment strategies among social animals. *Ethology* **112**: 729–747.

Komdeur, J. 2007. Constraints on evolutionary shifts in cooperative breeding. *Behav. Proc.* **76**: 75–77.

Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U. & Fehr, E. 2005. Oxytocin increases trust in humans. *Nature* **435**: 673–676.

Lamprecht, J. 1978. The relationship between food competition and foraging group size in some larger carnivores. A hypothesis. *Z. Tierpsychol.* **46**: 337–343.

Lehmann, L. & Keller, L. 2006. The evolution of cooperation and altruism – a general framework and a classification of models. *J. Evol. Biol.* **19**: 1365–1376.

Lehmann, L., Rousset, F., Roze, D. & Keller, L. 2007. Strong reciprocity or strong ferocity? A population genetic view of the evolution of altruistic punishment (p. 21). *Am. Nat.* **170**: 21–36.

Leimar, O. & Connor, R.C. 2003. By-product benefits, reciprocity, and pseudoreciprocity in mutualism. In: *Genetic and Cultural Evolution of Cooperation* (Hammerstein, P., ed.), pp. 203–222. MIT Press, Cambridge, MA.

Leimar, O. & Hammerstein, P. 2001. Evolution of cooperation through indirect reciprocity. *Proc. R. Soc. Lond. B Biol. Sci.* **268**: 745–753.

Lieberman, E., Hauert, C. & Nowak, M.A. 2005. Evolutionary dynamics on graphs. *Nature* **433**: 312–316.

Lotem, A., Fishman, M.A. & Stone, L. 2003. From reciprocity to unconditional altruism through signalling benefits. *Proc. R. Soc. Lond. B Biol. Sci.* **270**: 199–205.

Luce, R.D. & Raiffa, H. 1957. *Games and Decisions*. Wiley, New York.

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

Maynard Smith, J.M. & Price, G.R. 1973. Logic of animal conflict. *Nature* **246**: 15–18.

Mayr, E. 1961. Cause and effect in biology. *Science* **134**: 1501–1506.

McGregor, P.K. 1993. Signalling in territorial systems: a context for individual identification, ranging and eavesdropping. *Philos. Trans. R. Soc. Lond. B* **340**: 237–244.

McNamara, J.M., Barta, Z. & Houston, A.I. 2004. Variation in behaviour promotes cooperation in the prisoner's dilemma game. *Nature* **428**: 745–748.

Milinski, M. & Wedekind, C.1998. Working memory constrains human cooperation in the prisoner's dilemma. *Proc. Natl. Acad. Sci.* **95**: 13755–13758.

Noë, R. 2006. Cooperation experiments: coordination through communication vs. acting apart together. *Anim. Behav.* **71**: 1–18.

Nowak, M.A. 2006. Five rules for the evolution of cooperation. *Science* **314**: 1560–1563.

Nowak, M. & Sigmund, K. 1990. The evolution of stochastic strategies in the prisoners-dilemma. *Acta Appl. Math* **20**: 247–265.

Nowak, M.A. & Sigmund, K. 1992. Tit-for-tat in heterogeneous populations. *Nature* **355**: 250–253.

Nowak, M. & Sigmund, K. 1993. A strategy of win stay, lose shift that outperforms tit-for-tat in the prisoners-dilemma game. *Nature* **364**: 56–58.

Nowak, M.A. & Sigmund, K. 1998. Evolution of indirect reciprocity by image scoring. *Nature* **393**: 573–577.

Okasha, S. 2006. *Evolution and the Levels of Selection*. Oxford University Press, Oxford & New York.

Ostrom, E., Burger, J., Field, C.B., Norgaard, R.B. & Policansky, D. 1999. Sustainability – revisiting the commons: local lessons, global challenges. *Science* **284**: 278–282.

Pierce, N.E., Braby, M.F., Heath, A., Lohman, D.J., Mathew, J., Rand, D.B. & Travassos, M.A. 2002. The ecology and evolution of ant association in the Lycaenidae (Lepidoptera). *Annu. Rev. Entomol.* **47**: 733–771.

de Quervain, D.J.F., Fischbacher, U., Treyer, V., Schelthammer, M., Schnyder, U., Buck, A. & Fehr, E. 2004. The neural basis of altruistic punishment. *Science* **305**: 1254–1258.

Ratnieks, F.L.W. & Wenseleers, T. 2005. Policing insect societies. *Science* **307**: 54–56.

Roberts, G. 1998. Competitive altruism: from reciprocity to the handicap principle. *Proc. R. Soc. Lond. B Biol. Sci.* **265**: 427–431.

Roberts, G. 2005. Cooperation through interdependence. *Anim. Behav.* **70**: 901–908.

Roberts, G. & Sherratt, T.N. 2002. Behavioural evolution – does similarity breed cooperation? *Nature* **418**: 499–500.

Roberts, G. & Sherratt, T.N. 2007. Cooperative reading: some suggestions for integration of the cooperation literature. *Behav. Process.* **76**: 126–130.

Rutte, C. & Taborsky, M. 2007. Generalized Reciprocity in Rats. *PLoS Biol.* **5**: e196.

Sachs, J.L., Mueller, U.G., Wilcox, T.P. & Bull, J.J. 2004. The evolution of cooperation. *Q. Rev. Biol.* **79**: 135–160.

van Schaik, C.P. 1983. Why are diurnal primates living in groups? *Behaviour* **87**: 120–144.

Sherratt, T.N. & Roberts, G. 2002. The stability of cooperation involving variable investment. *J. Theor. Biol.* **215**: 47–56.

Sih, A., Bell, A. & Johnson, J.C. 2004. Behavioral syndromes: an ecological and evolutionary overview. *Trends Ecol. Evol.* **19**: 372–378.

Stephens, D.W., McLinn, C.M. & Stevens, J.R. 2002. Discounting and reciprocity in an iterated prisoner's dilemma. *Science* **298**: 2216–2218.

Sterck, E.H.M., Watts, D.P. & vanSchaik, C.P. 1997. The evolution of female social relationships in nonhuman primates. *Behav. Ecol. Sociobiol.* **41**: 291–309.

Tinbergen, N. 1963. On aims and methods in ethology. *Z. Tierpsychol.* **20**: 410–433.

Trivers, R.L. 1971. The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**: 35–57.

Wedekind, C. & Milinski, M. 2000. Cooperation through image scoring in humans. *Nature* **288**: 850–852.

West, S.A., Pen, I. & Griffin, A.S. 2002. Cooperation and competition between relatives. *Science* **296**: 72–75.

West, S.A., Gardner, A., Shuker, D.M., Reynolds, T., Burton-Chellow, M., Sykes, E.M., Guinnee, M.A. & Griffin, A.S. 2006a. Cooperation and the scale of competition in humans. *Curr. Biol.* **16**: 1103–1106.

West, S.A., Griffin, A.S., Gardner, A. & Diggle, S.P. 2006b. Social evolution theory for microbes. *Nat. Rev. Microbiol.* **4**: 597–607.

West, S.A., Griffin, A.S. & Gardner, A. 2007. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *J. Evol. Biol.* **20**: 415–432.

Wilson, D.S. 1980. *The Natural Selection of Populations and Communities*. Benjamin/Cummings, Menlo Park, CA.

Wilson, D.S. 1990. Weak altruism, strong group selection. *Oikos* **59**: 135–140.

Wilson, D.S. 2008. Social semantics: toward a genuine pluralism in the study of social behaviour. *J. Evol. Biol.* **21**: 368–373.

Wilson, D.S., Clark, A.B., Coleman, K. & Dearstyne, T. 1994. Shyness and boldness in humans and other animals. *Trends Ecol. Evol.* **9**: 442–446.

Wrangham, R.W. 1980. An ecological model of female-bonded primate groups. *Behaviour* **75**: 262–300.

Wynne, C.D.L. 2001. *Animal Cognition: The Mental Lives of Animals*. Palgrave, New York.

Zahavi, A. 1995. Altruism as a handicap – the limitations of kin selection and reciprocity. *J. Avian Biol.* **26**: 1–3.