Introduction

Cooperation and deception: from evolution to mechanisms

Sarah F. Brosnan¹,* and Redouan Bshary²

¹Department of Psychology, Georgia State University, PO Box 5010, Atlanta, GA 30302-5010, USA
²Department of Biology, University of Neuchâtel, Emile-Argand 11, Neuchâtel 2009, Switzerland

Nature is full of struggle, as predicted by the theory of evolution through natural selection, yet there are also paramount examples where individuals help each other. These instances of helping have been difficult to reconcile with Darwin’s theory because it is not always obvious how individuals are working for their own direct benefit. Consequently, initial publications that offered solutions to subsets of the observed cases of helping, such as kin selection or reciprocity, are among the most influential and most cited papers in evolution/behavioural ecology. During the last few years, a wave of new studies and concepts has considerably advanced our understanding of the conditions under which individuals are selected to help others. On the empirical side, advances are due to stronger incorporation of the natural history of each study species and an emphasis on proximate questions regarding decision-making processes. In parallel, theorists have provided more realistic models together with an increased exploration of the importance of life history and ecology in understanding cooperation. The ideas presented by the authors of this volume represent, in many ways, the revolutionary new approach to studying behaviour which is currently underway.

Keywords: cooperation; game theory; helping; deception; cheating; cognition

Nature is full of struggle, as predicted by the theory of evolution through natural selection, yet there are also paramount examples where individuals help each other. These instances of helping have been difficult to reconcile with Darwin’s theory because it is not always obvious how individuals are working for their own direct benefit. Consequently, initial publications that offered solutions to subsets of the observed cases of helping, such as kin selection (Hamilton 1964) or reciprocity (Trivers 1971; Axelrod & Hamilton 1981), are among the most influential and most cited papers in evolution/behavioural ecology. Despite these initial successes, models are often difficult to map onto actions, and empirical evidence for many proposed solutions is quite sparse. However, during the last few years, a wave of new studies and concepts has considerably advanced our understanding of the conditions under which individuals are selected to help others. We therefore think it is timely to bring together the state of the art concerning our knowledge of helping. Perhaps more importantly, reviewing our current knowledge should help us to identify the many gaps that still exist in our understanding of helping behaviour.

Of critical concern is that a large part of the existing theory is relatively poorly matched with empirical research. There may be several reasons for this mismatch. For example, theorists are often interested in human behaviour and the various forms of reciprocity that may lead to stable cooperation. As a consequence, very detailed models exist for these kinds of concepts. However, relatively few examples of cooperation and mutualism in other species seem to fit reciprocity concepts, as several contributors to this volume point out (Brosnan et al. 2010; Connor 2010; Leimar & Hammerstein 2010; Melis & Semmann 2010). Alternative concepts exist (Connor 2010; Leimar & Hammerstein 2010), but these seem to attract little attention from theorists. Another reason for the mismatch is that both ecological models and game theoretical models yield straightforward and seemingly simple cooperative solutions like ‘always help’ or ‘start cooperatively and then match your partner’s behaviour in subsequent rounds’, e.g. Tit for Tat; Axelrod & Hamilton (1981). Despite their simplicity and elegance, which lend a seductive allure, such predictions rarely fit real-life observations, where individuals do not seem to make decisions as precisely as predicted (de Waal & Suchak 2010) and where strong variation in behaviour is often observed within individuals, between individuals and between species (Soares et al. 2010). In fact, the effort to explain such variation has led to a major new field in behavioural ecology, named ‘animal personality’ (Gosling 2001; Sih et al. 2004; Bergmüller et al. 2010). Identifying ultimate and proximate sources of variation may help theorists to refine their models and hence...
make them more realistic (McNamara & Leimar 2010).

Overall, we feel that much important theory is already out there, but that more empirical studies are needed to help to identify the most important concepts. For example, predictions about how demography, life history and ecology affect the evolution and maintenance of helping are well established (Lehmann & Rousset 2010); what is lacking is empirical support. Thus, there is a clear need for careful descriptions of the natural history of species as a basis for well informed, and hence well designed, experiments. Such experiments should test not only the evolutionary theory, but also the mechanisms (cognitive, physiological, etc.) underlying cooperative behaviour. A detailed understanding of the mechanisms will provide key information for a new generation of more nuanced models. Mechanisms are important to specify trade-offs and constraints. For example, in most models there are a variety of typical assumptions, such as the assumptions that individuals have perfect memory, that gaining information is cost free, or that each individual is able to perform all different behavioural options investigated. These assumptions are unlikely to be realistic for any empirical example, and the conclusions that emerge from more realistic assumptions change the predictions of the models (McNamara & Leimar 2010).

Understanding the costs of information and associated trade-offs and constraints are particularly important for a full appreciation of how cooperating and avoiding cooperation (cheating) are—or are not—linked to the evolution of cognition and hence brain structure and brain size (Brosnan et al. 2010). Cooperating and cheating are closely intertwined, and it is unlikely that one can be understood in the absence of the other. This view emerges repeatedly in the contributions to this volume. Similarly, one can ask under which conditions stable cooperation is best achieved through ‘positive’ mechanisms that reward cooperative behaviour, versus when ‘negative’ mechanisms that handle non-cooperative partners through sanctioning, punishing or abandoning them may be the best option to promote helping. Negative mechanisms may easily promote cooperative behaviour by one partner but the outcome is not necessarily beneficial for all interacting individuals. Distinguishing sanctions and punishment of cheaters from coercion and parasitism remains a challenging topic for empiricists (Jensen 2010).

When we planned the content of this volume, our goal was to make sure that the general opinions voiced above were reflected in the contributions. Purely theoretical papers are in the minority; instead the emphasis was placed on empirical but nevertheless conceptual papers. More specifically, we felt that the major theoretical debates on group selection/cultural group selection versus inclusive fitness theory are well covered elsewhere, and that these positions are already clearly defined. We therefore avoid this topic in our volume, and focus on the ecological and game theoretical approaches instead. Lehmann & Rousset (2010) demonstrate how demography, life history and ecology may promote or hinder the evolution of helping by natural selection without explicitly distinguishing between direct and indirect fitness benefits or within-group and between-group competition. Similarly, de Waal & Suchak (2010) and Jaeggi et al. (2010) discuss helping (other-regarding) behaviour in other cooperative species, including cooperatively breeding species, without trying to split natural selection into these subunits. Thinking holistically in terms of natural selection may be one way around the theoretical discussions. Nevertheless, biologists have been trained in the inclusive fitness framework, and the relative importance of indirect versus direct fitness benefits (altruism versus cooperation) has been a key topic for empiricists. The realization that relatedness per se does not allow the conclusion that helping is due to an increase in indirect fitness (West et al. 2002; Lehmann & Rousset 2010) has led to an increased emphasis on direct benefits in cooperatively breeding species like meerkats (Clutton-Brock 2002), and the development of various concepts that strongly resemble or extend established game theoretic concepts of cooperation (Bergmüller et al. 2007). Measuring precisely the combined effects of direct and indirect fitness benefits of helping will be a major challenge for future studies. We refer interested readers to another recently edited Philosophical Transactions volume (Clutton-Brock et al. 2009) where contributors explored how direct and indirect fitness benefits interact in the formation of societies.

For the empirical chapters, we asked contributors to make their points using a broad taxonomic approach whenever feasible. Helping is widespread in nature, ranging from plants to insects, and from bacteria to humans. Examples from all major taxonomic groups exist, and comparing the evidence across a range of taxa may reveal important information about the generality or specificity of many of the concepts that are developed. Overall, we think that the various contributions provide major advances in understanding how cooperation, helping and deception actually manifest in nature and identify major future research areas with respect to four general issues. First, it is of paramount importance to study in detail (and then incorporate) the natural history of one’s study species. Second, we need to better integrate disciplines and research areas that currently focus on other topics into the study of cooperation. Third, we have to study mechanisms underlying behaviour and decision making. Finally, we need to better understand the degree to which helping—and also cheating and deception—are linked to the evolution of cognitive abilities, as well as the degree to which human cooperation may differ from that of other species. For further discussions of several important aspects related to this question we also refer interested readers to an edited Philosophical Transactions volume by Emery et al. (2007).

Before we summarize the major topics in this volume, we must deal briefly with terminology. Given the plethora of definitions for cooperation, altruism, and other terms in the field, confusion may emerge simply because of the ways different authors use the same terms, or use different terms for the same behaviour. In an effort to provide a coherent volume not only with respect to content but also
with respect to terminology, we asked our contributors to use our definitions for the most basic terms, derived from an evolutionary approach as described in Bshary & Bergmüller (2008), and to specify how any additional terms relate to these foundational definitions. By linking all of our papers under a common definitional framework, we provide readers a chance to see how ideas tie together in a way that is not possible when different definitions are used. The definitions summarized below cannot cover all aspects of helping, and authors were asked to provide definitions for additional terms or extensions of our terms in their chapters whenever necessary.

— **Helping**: this is the most general term and simply implies that an individual on average increases the fitness of a recipient. There are no assumptions about the costs or benefits to the helper.

— **Cooperative behaviour**: a behaviour that on average increases the fitness of a recipient and which is under positive selection if it on average increases the inclusive fitness of the actor via direct fitness benefits.

— **Altruistic behaviour**: a behaviour that on average increases the fitness of a recipient and which is under positive selection if it on average increases the inclusive fitness of the actor via indirect fitness benefits. This has also been called ‘biological altruism’. Some contributors distinguish this ‘ultimate altruism’ from what might be considered ‘proximate’ altruism, or, as it is often called, psychological altruism, which is defined by its underlying psychological mechanism, i.e. empathy. Psychological altruism does not ask how helping translates into fitness benefits.

— **Cooperation**: two (n) partners increase on average their direct fitness due to the interaction.

— **Cheating**: a behaviour that increases the immediate pay-off of the actor and reduces the immediate pay-off of the recipient. Cheating thus differs from deception, which implies a manipulation of the partner. Deception and cheating do not necessarily co-occur (although they may do so).

— **Spite**: a behaviour which decreases the direct fitness of both the actor and the recipient. Such behaviour may evolve if it increases the inclusive fitness of the actor via indirect fitness benefits. Similarly to altruistic behaviour, there must be a careful discrimination between ‘biological’ or ‘ultimate’ spite and ‘proximate’ spite. The latter is based on the motivation to hurt someone else, without asking the question how that may translate into fitness benefits.

— **Other-regarding behaviour**: This term has recently been used to describe helping behaviour, where the motivation to help is based on empathy rather than on calculations of how it might yield benefits to the actor.

Note that we find it essential to distinguish between individual behaviours and the outcome of interactions when we talk about cooperation (again, see Bshary & Bergmüller (2008) for more detail). Similarly, it is essential to distinguish ultimate function from proximate mechanisms; in both cases, mixing up the two can lead to misunderstandings. While we kept the definitions short, it is worthwhile to point out that West et al. (2007) included an important addition to the definition of cooperative behaviour (‘mutual benefits’ in their terminology), namely that the behaviour should in part be under selection because of its positive effect on the recipient. With this addition, one can exclude cases like elephants defecating and thereby providing by-product benefits to dung beetles that would otherwise fit the definition of cooperation.

### 1. THE RESULTS OF THE VOLUME

(a) **On the importance of knowing the natural history of one’s study species**

Only detailed knowledge about ecology and interaction patterns will allow informed guesses about the game structure in which individuals are engaged (n interactions, behavioural options, pay-off matrix, etc.) and, hence, how helping may increase the actors’ inclusive fitness. Demography, life history and ecology will be particularly important to understanding differences among species, but also among populations (Lehmann & Rousset 2010). The various game theoretic concepts presented primarily in Leimar & Hammerstein (2010) and Connor (2010), all implicitly make assumptions about the variables discussed in Lehmann & Rousset (2010) in order to construct pay-off matrices for the various behavioural options considered. Most importantly, however, models cannot make informed assumptions about trade-offs or constraints if these are not identified by empiricists. For example, the question whether interactions take place in front of bystanders and whether or not bystanders pay attention to these interactions is foremost an empirical question (Earley 2010), where the absence of any form of indirect reciprocity might be due to cognitive constraints or trade-offs between the benefits of acquiring information and the costs. One possibility is that the benefits of information collection will outweigh the costs only if inter-individual variation is high (McNamara & Leimar 2010).

A prime example of how ecology can be linked to the evolution of helping and to the evolution of specific mechanisms concerns reproductive systems. One emerging hypothesis is that cooperative breeders are more helpful than other, closely related, species due to the unusual constraints, and resulting interdependency, of their breeding system (Jaeggi et al. 2010). Under these circumstances, helping other group members becomes more unconditional on the recipients’ behaviour than might otherwise be the case. Indeed, one possibility which seems probable is that other-regarding (prosocial) motivations evolved (de Waal & Suchak 2010; Jaeggi et al. 2010) to provide a proximate mechanism that allows individuals to give away resources unconditionally in situations in which it is in their ultimate (if not immediate) benefit to do so.

(b) **Integrating new disciplines into the study of cooperation**

Three papers introduce disciplines that we believe are of major importance for an integrative

---

*Phil. Trans. R. Soc. B* (2010)
understanding of helping, but which currently focus
on other questions: animal personalities, behavioural
endocrinology, and communication network theory.
The observation that individuals consistently differ
with respect to behaviour has attracted great interest
in the last several years (Gosling 2001; Sih et al.
2004). The focus of research, however, has been on
the boldness—shyness axis and on aggression. Given
that inter-individual variation in behaviour may also
be a key factor promoting cooperation (McNamara &
Leimar 2010); Bergmüller et al. (2010) explore the
applicability of the concept of animal personalities to
cooperation. One important prediction is that it is
probable that cooperation emerges in context with
other behaviours (behavioural syndromes), and thus
that only a broad observational approach across con-
texts will yield a complete explanation of the
variation between individual levels of cooperation.

The paper by Soares et al. (2010) explores the
potential role of hormones in explaining differences
in levels of helping within individuals, between indi-
viduals and between species, linking endocrinology
to personality differences. Similarly to the field
of animal personality, behavioural endocrinology has
had a major focus on aggression, but may benefit
studies of cooperation. The paper also explains the
many potential pitfalls with respect to experimental
design and interpretation of results that one has to
take into account. Mechanisms can be studied on many
levels, be it ontogeny, physiology, endocrinology,
learning, cognition, processes in the brain,
and social environments, will help us better predict
and understand when cooperation occurs and when
it does not.

Finally, research on communication networks has
rarely focused on cooperation (McGregor 2005).
However, the fact that eavesdropping by bystanders
in communication networks has been documented in
a wide array of taxa, including invertebrates, makes it
probable that image scoring and behaviour adjust-
ments to being observed (‘audience effects’) may also
occur frequently in the context of cooperation.
Earley (2010) argues that the presence of bystanders
may provide a strong selective force on decisions to
cooperate, a force which is not currently considered
in an appropriate way.

(c) The need to study mechanisms underlying
behaviour and decision making
We expect that the study of mechanisms underlying
cooperative behaviour and decision-making processes
in general are most likely to impact our understanding
of cooperation in nature, especially in dialogue with
emerging theoretical models that take new evidence
into account. Mechanisms can be studied on many
different levels, be it ontogeny, physiology, endocrin-
ology, learning, cognition, processes in the brain,
genetics, interactions between genes and environment,
etc. It is clear that we are on the cusp of major advances
in understanding these mechanisms and their inter-
actions. New ideas for proximate mechanisms, at the
level of both causation and ontogeny, are emerging to
explain where cooperative behaviour does—and does
not—occur.

For instance, there has been a move away from the
idea of a calculated, precise model of reciprocal invest-
ment towards one based on rules of thumb and
emotional valence (Brosnan et al. 2010; de Waal &
Suchak 2010). This move opens the possibility of
explaining cooperation in a wider variety of species
and situations, and may provide explanations for
clearly cooperative interactions that nonetheless do
not meet traditional game theoretic rules (e.g. Tit for
Tat). Along with this, there is a greater interest in
how individuals develop cooperative behaviour. This
may emerge as a result of a genetically determined
strategy, a learned behavioural strategy (e.g. acquired
through associative learning), or higher cognition
(Brosnan et al. 2010). Of course, many potential
mechanisms may function at several of these levels
(e.g. empathy may be the result of genetic causation
linked with learning during ontogeny).

Moreover, although there are some notable
exceptions, in many cases we know little about the
decision-making processes underlying cooperation.
In many cases this may overlap with behavioural
mechanisms; for instance, although it is often assumed
that individuals make explicit calculations about their
decisions, there are remarkably few instances in
which this has been documented. Instead, it seems
probable that individuals are following rules of
thumb, including those based on emotional valence
and preferences for individuals (e.g. friendships;
Brosnan et al. 2010). Moreover, it is often unknown
what information individuals are using to make
decisions. They should be able to access their own
state, but may also be able to incorporate information from the environment or from others’
outcomes (e.g. social comparison; Brosnan & de
Waal 2003). Understanding how individuals make
decisions, and how they are affected by their physical
and social environments, will help us better predict
and understand when cooperation occurs and when
it does not.

(d) Linking cooperation, cheating and deception
to the evolution of cognitive abilities, and
implications for the ‘uniqueness’ of human
cooperation
A topic that is paramount in the current cooperation
literature and will without doubt remain hot for quite
some time is the question of how human cooperation
differs from that of other species (Melis & Semmann
2010). Although there are a range of assumptions
(Fehr Fischbacher & Gachter 2002; de Waal 2005),
the general consensus seems to be that we differ
from other species to some degree, but with clear evolu-
tionary continuities. In some ways, we might stand
apart further in the realm of deception and spite
than cooperation (Jensen 2010). The cause of those
differences is still a matter for debate.

One possibility that has been put forth is
that humans have a tendency to help others
(e.g. other-regarding preferences) which is beyond the
scope of other species (Silk et al. 2005; Jensen 2010).
On the other hand, evidence for these preferences has been found in other species, leading to two hypotheses, one more functional and the other more mechanistic (and thus, not mutually exclusive). First, it is possible that other-regarding preferences have evolved in cooperatively breeding species because their unique suite of life history characteristics leads to significant interdependency and thus increased opportunities for selection due to inclusive fitness (Jaeggi et al. 2010). Second, it may be that (one of) the proximate mechanism behind such preferences is empathy, which in at least some forms may be widespread in the animal kingdom (de Waal & Suchak 2010).

Other possibilities focus on the differences in magnitude or kind in other behaviours which in turn affect cooperation. For instance, although many other species show evidence of behavioural traditions, or cultures (e.g. Emery et al. 2007; Fragaszy & Perry 2003; Heyes & Galf 1996; Zentall & Galf 1988), human culture seems to be of a greater magnitude. Since culture is known to affect cooperation, even in individuals for whom basic criteria like age, level of education, etc. are similar (Gächter et al. 2010), humans may have an advantage in structuring interactions, relationships and institutions to favour cooperation. Humans also possess language, allowing them to communicate about other individuals more efficiently than other species are able to do (although see Connor 2010, for evidence in dolphins). This may lead to more efficient communication networks in humans, and increased opportunities for indirect learning, such as through reputations, than is available for other species (Earley 2010). For a more detailed discussion on the evolution of societies, including humans, we refer the reader to an edited volume by Clutton-Brock et al. (2009). Finally, the very destructiveness of humans’ behaviour may lead to increased cooperation, if the seeds of cooperation are sown through the hyper-competitiveness and spite seen in human societies (Jensen 2010).

2. CONCLUSION

It is clear that we are at an exciting time in the study of cooperation. After years of attempting to explain cooperation using fairly basic, dyadic models assuming static individuals (both developmentally within the same individual and across different individuals), we are beginning to understand the importance of variation at all levels in understanding cooperation. In large part this is the result of a move away from a reliance on the Prisoner’s Dilemma as the main concept to explain cooperative behaviour (Leimar & Hammerstein 2010), a concurrent acknowledgement that cooperation may involve more than two individuals (Connor 2010; Earley 2010), and, again, an increased recognition of the importance of life history and ecology in understanding cooperation (Lehmann & Rousset 2010; McNamara & Leimar 2010). As theory develops, it will be important for empiricists to follow with explicit tests of hypotheses and models, so that the theory can be further refined. The ideas presented by the authors of this volume represent, in many ways, the revolutionary new approach to studying behaviour which is currently underway.

We thank all of the authors who participated in this issue, all of whom enthusiastically embraced our challenge to think about new directions for the study of cooperation and generated a series of really stellar ideas. We also thank Manfred Milinski, as without his introduction this volume would not have occurred, and numerous peer reviewers, who further improved the volume. We also thank Claire Rawlinson at the Philosophical Transactions office for her assistance in putting this issue together. Funding to S.F.B. was provided by a National Science Foundation Human and Social Dynamics Grant (SES 0729244) and an NSF CAREER Award (SES 0847351), and by the Swiss Science Foundation to R.B.

REFERENCES


de Waal, F. B. M. 2005 Our inner ape: a leading primatologist explains why we are who we are. New York, NY: Riverhead.


